

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:33:46 ; Search time 70.07 Seconds
(without alignments)
711.845 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNLSKIENEYEV.....SNPLGISFLILMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 13230527 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:**
1: sp-archaea:
2: sp-bacteria:
3: sp-fungi:
4: sp-human:
5: sp-invertebrate:
6: sp-mammal:
7: sp-mhc:
8: sp-organelle:
9: sp-phase:
10: sp-plant:
11: sp-rodent:
12: sp-unclassified:
13: sp-vertebrate:
14: sp-virus:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1824.5	92.1	430	5 Q03999	Q03999 plasmodium
2	1824.5	92.1	509	5 Q25966	Q25966 plasmodium
3	1824.5	92.1	539	5 Q25976	Q25976 plasmodium
4	1824.5	92.1	539	5 Q25984	Q25984 plasmodium
5	1824.5	92.1	539	5 Q9TYG1	Q9TYG1 plasmodium
6	1821.5	91.9	539	5 Q25972	Q25972 plasmodium
7	1821.5	91.9	539	5 Q25981	Q25981 plasmodium
8	1818.5	91.8	539	5 Q25971	Q25971 plasmodium
9	1818.5	91.8	539	5 Q25973	Q25973 plasmodium
10	1167.5	58.9	651	5 Q25924	Q25924 plasmodium
11	1164.5	58.8	569	5 Q25969	Q25969 plasmodium
12	1164.5	58.8	569	5 Q25974	Q25974 plasmodium
13	1164.5	58.8	569	5 Q25975	Q25975 plasmodium
14	1164.5	58.8	569	5 Q25977	Q25977 plasmodium
15	1164.5	58.8	569	5 Q25979	Q25979 plasmodium
16	1164.5	58.8	1694	5 Q9TZT5	Q9TZT5 plasmodium
17	1164.5	58.8	1694	5 Q9NHX1	Q9NHX1 plasmodium
18	1164.5	58.8	1704	5 Q9TZT4	Q9TZT4 plasmodium
19	1163.5	58.7	569	5 Q25983	Q25983 plasmodium

20	1151.5	58.1	569	5	Q25967	Q25967 plasmodium
21	1145.5	57.8	569	5	Q25970	Q25970 plasmodium
22	1145.5	57.8	569	5	Q25980	Q25980 plasmodium
23	1145.5	57.8	569	5	Q25982	Q25982 plasmodium
24	1145	57.8	570	5	Q25968	Q25968 plasmodium
25	1145	57.8	570	5	Q9TYG2	Q9TYG2 plasmodium
26	1144.5	57.7	569	5	Q25978	Q25978 plasmodium
27	1134	57.2	1720	5	Q25922	Q25922 plasmodium
28	1130	57.0	652	5	Q25923	Q25923 plasmodium
29	1069	53.9	373	5	Q25724	Q25724 plasmodium
30	1067.5	53.9	372	5	Q25725	Q25725 plasmodium
31	1066	53.8	373	5	Q25721	Q25721 plasmodium
32	1063.5	53.7	372	5	Q43997	Q43997 plasmodium
33	1063	53.6	373	5	Q43995	Q43995 plasmodium
34	1062.5	53.6	372	5	Q25726	Q25726 plasmodium
35	1059	53.4	373	5	Q25722	Q25722 plasmodium
36	1058	53.4	373	5	Q25723	Q25723 plasmodium
37	1048.5	52.9	372	5	Q25719	Q25719 plasmodium
38	1048.5	52.9	372	5	Q25720	Q25720 plasmodium
39	1048	52.9	373	5	Q25727	Q25727 plasmodium
40	1047.5	52.9	372	5	Q25717	Q25717 plasmodium
41	1045	52.7	373	5	Q43996	Q43996 plasmodium
42	1043	52.6	373	5	Q25728	Q25728 plasmodium
43	1042.5	52.6	372	5	Q25718	Q25718 plasmodium
44	1024.5	51.7	219	5	Q9UB87	Q9UB87 plasmodium
45	747	37.7	599	5	Q9NCN2	Q9NCN2 plasmodium

ALIGNMENTS

RESULT 1
Q03999 ID Q03999 PRELIMINARY; PRT; 400 AA.
AC Q03999;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE MEROZOITE SURFACE ANTIGEN PRECURSOR 1 (PMMSA) (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RL Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: COULD BE ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -!- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC EMBL; M64681; AAA29709.1; -;
DR InterPro; IPR000561; -;
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT NON_TER 1
FT SIGNAL <1 24 POTENTIAL.
FT CHAIN 25 400 83K MEROZOITE SURFACE ANTIGEN.
FT CHAIN 25 286 42K MEROZOITE SURFACE ANTIGEN.
FT CHAIN 287 400 19K MEROZOITE SURFACE ANTIGEN.
FT TRANSMEM 383 400 MEMBRANE ANCHOR.
SQ SEQUENCE 400 AA; 45824 MW; 537F075058626AC2 CRC64;

Query Match 92.1%; Score 1824.5; DB 5; Length 400;

Best Local Similarity 93.6%;

Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

OY 1 AVTTSVIDNLSKIENEYEVLYKPLAGVYRSKQLQENNVMTFNVNVKDLNLSRFNKR 60

```
Db 25 AVTPSVIDNLSKIENEYEVLYKPLAGYRSLKKOLENNVMTFNVNVDILNSRFNKRE 84
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDKFLSSYNYIKDSITDITINFANDVL 120
Db 85 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDKFLSSYNYIKDSITDITINFANDVL 144
QY 121 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFLNNIETLYKTVNDKIDLFVHLEAKV 180
Db 145 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFLNNIETLYKTVNDKIDLFVHLEAKV 204
QY 181 LNYTEKSNVEVKIKELNYLKTIOKLADFKKNNFVGIADLSTDYNNHNLTKFLSTGM 240
Db 205 LNYTEKSNVEVKIKELNYLKTIOKLADFKKNNFVGIADLSTDYNNHNLTKFLSTGM 264
QY 241 VFENLLKSVLSNLLDWMKLARYVHKFTTPMRKKTMTQQSSGCFRHLDEREECKCLLNYKOE 300
Db 265 VFENLAKTVLSNLLDGNL-OQMLNISQHCVKKQCPQNSGCFRHLDEREECKCLLNYKOE 323
QY 301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGNGKKITCECTKPCDYPFLDGFICSSSNF 360
Db 324 GDKCVENPNPTCNENNGGCDADAKCTEEDSGNGKKITCECTKPCDYPFLDGFICSSSNF 383
QY 361 LGISFLLIIMLILYSFI 377
Db 384 LGISFLLIIMLILYSFI 400
```

```
RESULT 2
Q25966
ID Q25966 PRELIMINARY; PRT; 539 AA.
AC Q25966;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSP1) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13357; BAA02618.1; -
DR InterPro; IPR000561; -
DR InterPro; IPR001245; -
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61144 MW; 7B7DE90C1D0ACDE7 CRC64;
```

```
Query Match 92.1%; Score 1824.5; DB 5; Length 539;
Best Local Similarity 93.6%; Pred. No. 2e-87;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;
```

```
QY 1 AVTTSVIDNLSKIENEYEVLYKPLAGYRSLKKOLENNVMTFNVNVDILNSRFNKRE 60
Db 164 AVTPSVIDNLSKIENEYEVLYKPLAGYRSLKKOLENNVMTFNVNVDILNSRFNKRE 223
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDKFLSSYNYIKDSITDITINFANDVL 120
Db 224 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDKFLSSYNYIKDSITDITINFANDVL 283
QY 121 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFLNNIETLYKTVNDKIDLFVHLEAKV 180
Db 284 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFLNNIETLYKTVNDKIDLFVHLEAKV 343
```

```
QY 181 LNYTEKSNVEVKIKELNYLKTIOKLADFKKNNFVGIADLSTDYNNHNLTKFLSTGM 240
Db 344 LNYTEKSNVEVKIKELNYLKTIOKLADFKKNNFVGIADLSTDYNNHNLTKFLSTGM 403
QY 241 VFENLLKSVLSNLLDWMKLARYVHKFTTPMRKKTMTQQSSGCFRHLDEREECKCLLNYKOE 300
Db 404 VFENLAKTVLSNLLDGNL-OQMLNISQHCVKKQCPQNSGCFRHLDEREECKCLLNYKOE 462
QY 301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGNGKKITCECTKPCDYPFLDGFICSSSNF 360
Db 463 GDKCVENPNPTCNENNGGCDADAKCTEEDSGNGKKITCECTKPCDYPFLDGFICSSSNF 522
QY 361 LGISFLLIIMLILYSFI 377
Db 523 LGISFLLIIMLILYSFI 539
```

```
RESULT 3
Q25976
ID Q25976 PRELIMINARY; PRT; 539 AA.
AC Q25976;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSP1) of Plasmodium falciparum from
field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13362; BAA02623.1; -
DR InterPro; IPR000561; -
DR InterPro; IPR001245; -
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61045 MW; 8BF9C64322E9A778 CRC64;
```

```
Query Match 92.1%; Score 1824.5; DB 5; Length 539;
Best Local Similarity 93.6%; Pred. No. 2e-87;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;
```

```
QY 1 AVTTSVIDNLSKIENEYEVLYKPLAGYRSLKKOLENNVMTFNVNVDILNSRFNKRE 60
Db 164 AVTPSVIDNLSKIENEYEVLYKPLAGYRSLKKOLENNVMTFNVNVDILNSRFNKRE 223
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDKFLSSYNYIKDSITDITINFANDVL 120
Db 224 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDKFLSSYNYIKDSITDITINFANDVL 283
QY 121 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFLNNIETLYKTVNDKIDLFVHLEAKV 180
Db 284 GYKILSEKYSKSDLSIKKYINDKOGENEKYLPLFLNNIETLYKTVNDKIDLFVHLEAKV 343
QY 181 LNYTEKSNVEVKIKELNYLKTIOKLADFKKNNFVGIADLSTDYNNHNLTKFLSTGM 240
Db 344 LNYTEKSNVEVKIKELNYLKTIOKLADFKKNNFVGIADLSTDYNNHNLTKFLSTGM 403
QY 241 VFENLLKSVLSNLLDWMKLARYVHKFTTPMRKKTMTQQSSGCFRHLDEREECKCLLNYKOE 300
Db 404 VFENLAKTVLSNLLDGNL-OQMLNISQHCVKKQCPQNSGCFRHLDEREECKCLLNYKOE 462
QY 301 GDKCVENPNPTCNENNGGCDADAKCTEEDSGNGKKITCECTKPCDYPFLDGFICSSSNF 360
```

```
|||||
Db 463 GDKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNF 522
QY 361 LGISFLLILMLILYSFI 377
Db 523 LGISFLLILMLILYSFI 539

RESULT 4
Q25984
ID Q25984 PRELIMINARY; PRT: 539 AA.
AC Q25984;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13356; BAA02617.1; -
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61114 MW; 3788015F3127CB9E CRC64;
```

```
Query Match 92.1%; Score 1824.5; DB 5; Length 539;
Best Local Similarity 93.6%; Pred. No. 2e-87;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEYEVLYLKLPLAGYVRSLLKQLENNVMTFNVNKKDILNSRFNKR 60
Db 164 AVTSPVIDNLSKIENEYEVLYLKLPLAGYVRSLLKQLENNVMTFNVNKKDILNSRFNKR 223
QY 61 NFKNVLESLLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSNYIKDSTDTDFINFANDVL 120
Db 224 NFKNVLESLLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSNYIKDSTDTDFINFANDVL 283
QY 121 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDNKTDLFVHLEAKV 180
Db 284 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDNKTDLFVHLEAKV 343
QY 181 LNYTSEKSNVEVKIKELNYLKTIDQKLADFKNKNNFVGIADLSTDYNHNNLLTKFLSTGM 240
Db 344 LNYTSEKSNVEVKIKELNYLKTIDQKLADFKNKNNFVGIADLSTDYNHNNLLTKFLSTGM 403
QY 241 VFENLLKSVLSNLLDQNLARYVHFTTPMRKMTMQSSGCGFRHLDERECKCLLNKYOE 300
Db 404 VFENLAKTVLSNLLDQNL-OGMLNISQHCYVKQCPQNSGCGFRHLDERECKCLLNKYOE 462
QY 301 GDKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNF 360
Db 463 GDKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNF 522
QY 361 LGISFLLILMLILYSFI 377
Db 523 LGISFLLILMLILYSFI 539

RESULT 5
Q25972
ID Q25972 PRELIMINARY; PRT: 539 AA.
AC Q25972;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
```

```
ID Q9TYG1 PRELIMINARY; PRT: 539 AA.
AC Q9TYG1;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPI) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13358; BAA02619.1; -
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61075 MW; C70C2E100EC4A101 CRC64;
```

```
Query Match 92.1%; Score 1824.5; DB 5; Length 539;
Best Local Similarity 93.6%; Pred. No. 2e-87;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEYEVLYLKLPLAGYVRSLLKQLENNVMTFNVNKKDILNSRFNKR 60
Db 164 AVTSPVIDNLSKIENEYEVLYLKLPLAGYVRSLLKQLENNVMTFNVNKKDILNSRFNKR 223
QY 61 NFKNVLESLLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSNYIKDSTDTDFINFANDVL 120
Db 224 NFKNVLESLLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSNYIKDSTDTDFINFANDVL 283
QY 121 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDNKTDLFVHLEAKV 180
Db 284 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDNKTDLFVHLEAKV 343
QY 181 LNYTSEKSNVEVKIKELNYLKTIDQKLADFKNKNNFVGIADLSTDYNHNNLLTKFLSTGM 240
Db 344 LNYTSEKSNVEVKIKELNYLKTIDQKLADFKNKNNFVGIADLSTDYNHNNLLTKFLSTGM 403
QY 241 VFENLLKSVLSNLLDQNLARYVHFTTPMRKMTMQSSGCGFRHLDERECKCLLNKYOE 300
Db 404 VFENLAKTVLSNLLDQNL-OGMLNISQHCYVKQCPQNSGCGFRHLDERECKCLLNKYOE 462
QY 301 GDKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNF 360
Db 463 GDKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPDSPYPLFDGIFCSSNF 522
QY 361 LGISFLLILMLILYSFI 377
Db 523 LGISFLLILMLILYSFI 539

RESULT 6
Q25972
ID Q25972 PRELIMINARY; PRT: 539 AA.
AC Q25972;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
```


[illegible]

Query Match	91.8%;	Score 1818.5;	DB 5;	Length 539;
Best Local Similarity	93.4%;	pred. No. 4.1e-87;		
Matches 352;	Conservative 6;	Mismatches 18;	Indels 1;	Gaps 1;

Qy	1	AVTSPVIDNILSKTENEVEVLYLPLAGVYRSLLKKQLENNVMTFNNVVKDILNSRFNKRRE	60
Db	164	AVTSPVIDNILSKTENEVEVLYLPLAGVYRSLLKKQLENNVMTFNNVVKDILNSRFNKRRE	223

Qy	61	NFKWVLESDDLIPYKDLTSSNYVVDPPYKFLNKRKDKFLSSVNYTKDSITDITDFANDVL	120
Db	224	NFKWVLESDDLIPYKDLTSSNYVVDPPYKFLNKRKDKFLSSVNYTKDSITDITDFANDVL	283

Qy	121	GYKILSEKYSKLDLSIKKYNINOKQGENEKLPLNNIETLYKTVDNKIDLFVHLEAKV	180
Db	284	GYKILSEKYSKLDLSIKKYNINOKQGENEKLPLNNIETLYKTVDNKIDLFVHLEAKV	343
Qy	181	LNITYEKSNNVEVKIKELNYLTKTQDKLADFKKNNFVGIADLSTDYNNHNNLLTKFLSTGM	240
Db	344	LNITYEKSNNVEVKIKELNYLTKTQDKLADFKKNNFVGIADLSTDYNNHNNLLTKFLSTGM	403
Qy	241	VFENLLKSVLSNLLDWKLARYKHVHTPPMRKKTMIQOSSGCFRHLDERECKCLLNYKOE	300
Db	404	VFENLAKTVLSNLLDGNL-QCMLNISQHCQVKKQCPONSGCFRHLDERECKCLLNYKOE	462
Qy	301	GDKCVENPNPTCENNNGCCDADAKCTEEDSGSNKKITCECTKPDYPLFDGIFCSSSNF	360
Db	463	GSKCVENPNPTCENNNGCCDADAKCTEEDSGSNKKITCECTKPDYPLFDGIFCSSSNF	522
Qy	361	LGISFLLILMLILYSFI 377	
Db	523	LGISFLLILMLILYSFI 539	
RESULT 10			
ID	Q25924	PRELIMINARY; PRT; 651 AA.	
AC	Q25924		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DE	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)		
DE	MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=RO-71;		
RX	MEDLINE=52275047; PubMed=1592091;		
RA	Alafsson P., Matile H., Certa U.;		
RT	"Plasmodium falciparum: the repetitive MSA-1 surface protein of the		
RT	RO-71 isolate is recognized by mouse antibody against the		
RT	nonrepetitive repeat block of RO-33."		
RL	Exp. Parasitol. 74:381-389(1992).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=RO-71;		
RX	MEDLINE=95354793; PubMed=7628566;		
RA	Tolle R., Bujard H., Cooper J.A.;		
RT	"Plasmodium falciparum: variations within the C-terminal region of		
RT	merozoite surface antigen-1."		
RL	Exp. Parasitol. 81:47-54(1995).		
DR	ENBL; 235329; CAA84558.1;		
DR	InterPro: IPR000561;		
DR	Pfam; PF00008; EGF; 1.		
KW	Merozoite.		
FT	NON_TER		
SQ	SEQUENCE 651 AA; 74134 MW; AA2137B699255150 CRC64;		

Query Match	58.9%;	Score 1167.5;	DB 5;	Length 651;
Best Local Similarity	57.9%;	Pred. No. 2.6e-53;		
Matches 226;	Conservative	59;	Mismatches 84;	Indels 21;
				Gaps 3;

Qy	7	IDNLSKIENEYEVYLLKPLAGVYRSLKOLENNVMTFNVNVKDILNSRFNKRENFKNVL	66
		: : : : : : : : : : : : :	
Db	264	MDNLSGFENEYDVIYLLKPLAGVYRSLLKQIEKNIFNLNDILNSRLKRRYFLDVL	323
Qy	67	ESDLIPYKDLTSSWYVYKDPYKFLNKKRDKFLSSYNYIKDSIDTDINFANDWLVGYKIL	126
		: : : : : : : : : : : : :	
Db	324	ESDLMQFKHISSENEYIIEDSFKLNSQKNTLLKSRYIKESVENDIKFAQEGISYYEKV	383
Qy	127	SEKYKSDLSTKKYI-----NDKQGENEYKLPFLNNIETLYKTVND	167
		: : : : : : : : : : : : :	
Db	384	LAKYKDDLESITKKVKEKEFFSPSPPTPPSPAKTDQSKESFLPPLTMYLYNNLVN	443


```
QY 7 IDNILSKIENEYEVLYKPLAGVYRSKKOLENNVMTFNVVVKDILNSRFNKRENFKNVL 66
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
182 MNILSGFENEVDVYLYKPLAGVYRSKKOLENNVMTFNVVVKDILNSRFNKRENFKNVL 241
QY 67 ESDLIPYKDLTSSNVVVKDYPYKFLAKRKDFLSSVYIKDSIDTDINFANDVLGYKIL 126
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
242 ESDLMOFKHISSNEYTIEDSFLLNSEQNTLLSKYKIKESVENDIKFAQEGISYYEKV 301
QY 127 SEKYKSDLDSIKKIYI-----NDKQGENEKYLPFLNNIETLYKTVD 167
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
302 LAKYKDDLESIRKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNLVN 361
QY 168 KIDLFIHLEAKVLYNVEKSNVEVKIKELNYLKTIOCKLADFKKNNFVGIADLSTDYN 227
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
362 KIDDLINLAKINDONVEKDEAHVKITKLSDLKADDDKIDLFKNTNDFEAIKKLINDDT 421
QY 228 HNNLTFLSTGWFENLKSVLNSLLDWKLARYVKHFTTPMRKKTMIQSSGCFRHLDE 287
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
422 KKDMLGKLLSTGLV-QNFPNTIISKIEGKF-QDMLNISQHCVCVKQCPENSGCFRHLDE 479
QY 288 REECKLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDY 347
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
480 REECKLLNYKQEGDKCVENPNPTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDY 539
QY 348 PLFDGIFCSSNPLGISFLLILMLIYSPF 377
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
540 PLFDGIFCSSNPLGISFLLILMLIYSPF 569
```

Search completed: August 8, 2001, 12:33:47
Job time: 275 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:32:38 ; Search time 24.79 Seconds
(without alignments)
520.949 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982

Sequence: 1 AVTTSVIDNLSKIENEYEV.....SNFLGISFLLMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1824.5	92.1	1630	1 MSP1_PLAPK	P04932 plasmodium
2	1824.5	92.1	1639	1 MSP1_PLAPW	P04933 plasmodium
3	1164	58.7	1726	1 MSP1_PLAPC	P04934 plasmodium
4	1163	58.7	1726	1 MSP1_PLAPP	P50495 plasmodium
5	1156.5	58.4	1682	1 MSP1_PLAPF3	P19598 plasmodium
6	1145	57.8	1701	1 MSP1_PLAPF	P13819 plasmodium
7	1145	57.8	1701	1 MSP1_PLAPM	P08569 plasmodium
8	588.5	29.7	1772	1 MSP1_PLAYO	P13828 plasmodium
9	142	7.2	1251	1 RBP2_PLAVB	Q00799 plasmodium
10	139.5	7.0	377	1 Y704_METJA	Q58115 methanococc
11	136.5	6.9	2198	1 YLJ2_CAEEL	P34367 caenorhabdi
12	135	6.8	1102	1 RPOP_AGABT	P33539 agaricus bi
13	134	6.8	1150	1 IRL1_YEAST	P40541 saccharomyc
14	133	6.7	748	1 Y875_METJA	Q38285 methanococc
15	133	6.7	864	1 CHEA_BORBU	Q44737 borrelia bu
16	133	6.7	1030	1 Y018_MYCPN	P75093 mycoplasma
17	130.5	6.6	881	1 LHS1_YEAST	P36016 saccharomyc
18	130	6.6	1162	1 BXCN_CLOBO	Q06366 clostridium
19	130	6.6	1196	1 BXCN_CLOBO	P46081 clostridium
20	129.5	6.5	1956	1 ATX1_PLAFA	Q04956 plasmodium
21	129	6.5	756	1 Y328_MYCGE	Q49419 mycoplasma
22	129	6.5	1162	1 BXEN_CLOBO	P46082 clostridium
23	129	6.5	1169	1 EX5B_BORBU	O51578 borrelia bu
24	128.5	6.5	867	1 DPOL_RICPR	Q05949 rickettsia
25	127	6.4	766	1 Y583_CAEEL	Q09622 caenorhabdi
26	126.5	6.4	944	1 NUF1_YEAST	P32380 saccharomyc
27	126	6.4	703	1 YH83_PHANI	P51819 pharbitis n
28	125.5	6.3	655	1 YKDA_MYCPA	P45615 mycoplasma
29	125	6.3	1038	1 CIN8_YEAST	P27895 saccharomyc
30	125	6.3	1901	1 YCF1_TOBAC	P12222 nicotiana t
31	124.5	6.3	2869	1 RBP1_PLAVB	Q00798 plasmodium
32	123.5	6.2	989	1 SERA_PLAPG	P13823 plasmodium
33	122.5	6.2	749	1 MAD1_YEAST	P40957 saccharomyc

ALIGNMENTS

RESULT 1

ID	MSP1_PLAPK	STANDARD	PRT	1630 AA
AC	P04932:			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS) (PMMSA) (P190).			
DE	MSP-1.			
GN	Plasmodium falciparum (isolate K1 / Thailand).			
OS	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OC	NCBI_TaxID=5839;			
OX	(1)			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=86136024; PubMed=3004972;			
RX	Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,			
RA	Stunnenberg H., Bujard H.;			
RT	"Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";			
RL	EMBO J. 4:3823-3829(1985).			
RN	(2)			
RP	REVISTONS. SEQUENCE FROM N.A.			
RA	Pan W., Tolle R., Bujard H.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).			
CC	-1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
EMBL	X03371; CAA27070.1; -			
DR	P1; A25120; SAZOK1.			
DR	InterPro; IPR000561; -			
DR	Pfam; PF00008; EGF; 1.			
KW	Malaria; Merozoite; GPI-anchored; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.			
FT	SIGNAL 1 19			POTENTIAL.
FT	CHAIN 20 1630			MEROZOITE SURFACE PROTEIN 1.
FT	DOMAIN 67 84			TRIPEPTIDE SG(TP) REPEAT.
FT	TRANSMEM 1614 1630			MEMBRANE ANCHOR.
FT	CARBOHYD 97 97			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 259 259			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 755 755			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 759 759			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 774 774			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 835 835			N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD 911 911			N-LINKED (GLCNAC. .) (POTENTIAL).

Q08372 plasmodium
P75440 mycoplasma
P39682 saccharomyc
O51246 borrelia bu
P19214 plasmodium
P25386 saccharomyc
P39526 saccharomyc
Q14525 homo sapien
Q10310 schizosacch
P53959 saccharomyc
Q05663 mesocricetu
P05475 kluyveromyc

34 122 6.2 3135 1 S230_PLAFO
35 121 6.1 632 1 Y242_MYCPN
36 120.5 6.1 629 1 PR39_YEAST
37 119.5 6.0 971 1 Y228_BORBU
38 119.5 6.0 1435 1 EBAL_PLAPC
39 119.5 6.0 1790 1 USOL_YEAST
40 119.5 6.0 2014 1 YJ07_YEAST
41 119 6.0 404 1 K1MB_HUMAN
42 118.5 6.0 515 1 YD56_SCHPO
43 118.5 6.0 839 1 YNE1_YEAST
44 118.5 6.0 845 1 SCPL1_MESAU
45 118 6.0 453 1 YKP9_KLULA

```
FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;

Query Match 92.1%; Score 1824.5; DB 1; Length 1630;
Best Local Similarity 93.6%; Pred. No. 2.5e-90;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEVEVLKPLAGVYRSLKQLENNVMTFNVNVDILNSRFNKR 60
Db 1255 AVTSPVIDNLSKIENEVEVLKPLAGVYRSLKQLENNVMTFNVNVDILNSRFNKR 1314

QY 61 NFKNVLESLLIPYKDLTSSNVVVDKPYKFLNKRKDFLSSYNYIKDSIDTDINFANDVL 120
Db 1315 NFKNVLESLLIPYKDLTSSNVVVDKPYKFLNKRKDFLSSYNYIKDSIDTDINFANDVL 1374

QY 121 GYKILSEKYSKDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKLIDLFVHLEAKV 180
Db 1375 GYKILSEKYSKDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKLIDLFVHLEAKV 1434

QY 181 LNYTYEKSNEVKIKELNYLKTIOQKLDADFKNNFVGTADLSTDYNNHLLTKFLSTGM 240
Db 1435 LNYTYEKSNEVKIKELNYLKTIOQKLDADFKNNFVGTADLSTDYNNHLLTKFLSTGM 1494

QY 241 VFENLLKSVLSNLLDWKLARYVHKFTTPMRKKTMTIQSSGCFRHLDERECKCLNYKOE 300
Db 1495 VFENLAKTVLSNLLDGNL-QGMLNISQHCYKQCPQNSGCFRHLDERECKCLNYKOE 1553

QY 301 GDKVCENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPFLDGFICSSSNF 360
Db 1554 GDKVCENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPFLDGFICSSSNF 1613

QY 361 LGISFLILMLILYSFI 377
Db 1614 LGISFLILMLILYSFI 1630

RESULT 2
MSPI_PLAFW
ID MSPI_PLAFW STANDARD; PRT; 1639 AA.
AC P04933.
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A.; Lockyer M.J.; Odink K.G.; Sandhu J.S.; Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites."
RL Nature 317:270-273(1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02919; CAA26676.1; -
DR PIR; A24594; A24594.
DR InterPro; IPR000561; -
DR Pfam; PF000008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
FT SIGNAL 1 19
FT CHAIN 20 1639
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 768 768 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 92.1%; Score 1824.5; DB 1; Length 1639;
Best Local Similarity 93.6%; Pred. No. 2.6e-90;
Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEVEVLKPLAGVYRSLKQLENNVMTFNVNVDILNSRFNKR 60
Db 1264 AVTSPVIDNLSKIENEVEVLKPLAGVYRSLKQLENNVMTFNVNVDILNSRFNKR 1323

QY 61 NFKNVLESLLIPYKDLTSSNVVVDKPYKFLNKRKDFLSSYNYIKDSIDTDINFANDVL 120
Db 1324 NFKNVLESLLIPYKDLTSSNVVVDKPYKFLNKRKDFLSSYNYIKDSIDTDINFANDVL 1383

QY 121 GYKILSEKYSKDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKLIDLFVHLEAKV 180
Db 1384 GYKILSEKYSKDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKLIDLFVHLEAKV 1443

QY 181 LNYTYEKSNEVKIKELNYLKTIOQKLDADFKNNFVGTADLSTDYNNHLLTKFLSTGM 240
Db 1444 LNYTYEKSNEVKIKELNYLKTIOQKLDADFKNNFVGTADLSTDYNNHLLTKFLSTGM 1503

QY 241 VFENLLKSVLSNLLDWKLARYVHKFTTPMRKKTMTIQSSGCFRHLDERECKCLNYKOE 300
Db 1504 VFENLAKTVLSNLLDGNL-QGMLNISQHCYKQCPQNSGCFRHLDERECKCLNYKOE 1562

QY 301 GDKVCENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPFLDGFICSSSNF 360
Db 1563 GDKVCENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPFLDGFICSSSNF 1622

QY 361 LGISFLILMLILYSFI 377
Db 1623 LGISFLILMLILYSFI 1639

RESULT 3
MSPI_PLAFW
ID MSPI_PLAFW STANDARD; PRT; 1726 AA.
AC P04934.
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
```

DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5835;
 RN [1]
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE=86205236; PubMed=3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A., Wolff R.;
 RT "Variation in the gene encoding a major merozoite surface antigen of
 the human malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 14:3311-3323(1986).
 RN [2]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=88143999; PubMed=3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X03831; CAZ27446.1; -.
 DR PIR: A23386; SAAZQM.
 DR InterPro: IPR000561; -.
 DR Pfam: PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196197 MW; DD8AD45FA352BCF3 CRC64;

Query Match 58.7%; Score 1164; DB 1; Length 1726;
 Best Local Similarity 57.5%; Pred. No. 4.7e-55;
 Matches 225; Conservative 60; Mismatches 84; Indels 22; Gaps 3;
 QY 7 IDNLSKTENEVEVLYLPLAGVYSLKOLNNVMTFNVNVKDLNLSRKNREKFNVL 66
 DB 1338 MDNLSGGENEYVLYLPLAGVYSLKQIEKNFTFNLDNLSLKRKRKYFLDVL 1397
 QY 67 ESDLIPYKDLTSSNVKDPYKFLNKRKDRKFLSSYNTIKDSIDTDFNFANDVLGYKIL 126
 DB 1398 ESDLQMFKHSSNEVIEDSPKLLSEKNTLLKYYKIKESVENDIKFAQEGISYEKV 1457
 QY 127 SKYKSDLSKTKYI-----NDKQGENEYKLPFLNNIETLYKTVN 166
 DB 1458 LAKYKDDLESIKVKEBKEKPPSPPTTPPSPAKTDBQKESKFLPFLNITLYNNLV 1517

QY 167 DKIDLFVHLEAKVLNYYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDY 226
 DB 1518 NKIDDYLNLRKAKINDCNVDEAHVKITKLSDAIKDDKIDLFKNHNDFAIKKLINDD 1577
 QY 227 NHNNLLTKFLSTGMVFNLLKSVLSNLNDWLKARYVYKFTTPMRKKTWIOSSGCGFRHLD 286
 DB 1578 TKKMDLKGKLLSTGLV-QNFPNTIISKLIETGKFP-QDMLNISQHCQVKKQCPENSGCGRHLD 1635
 QY 267 ERECKCLLNKQEGDKCVENPNTCNENNGCGDADAKTEDSGSNKKITCECTKPD 346
 DB 1636 ERECKCLLNKQEGDKCVENPNTCNENNGCGDADAKTEDSGSNKKITCECTKPD 1695
 QY 347 YPLFDGIFCSSNFGISFLILMLILYSFI 377
 DB 1696 YPLFDGIFCSSNFGISFLILMLILYSFI 1726
 RESULT 4
 MSP1_PLAPP
 ID MSP1_PLAPP STANDARD; PRT; 1726 AA.
 AC P50495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (GP195).
 GN MSP-1
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Keto A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydropathy profile of the
 major merozoite surface antigen (gp195) of the Uganda-Palo Alto
 isolate.";
 RL Exp. Parasitol. 67:1-11(1988).
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M37213; AAA29611.1; -.
 DR InterPro: IPR000561; -.
 DR Pfam: PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).

```
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEFA2F9A026 CRC64;

Query Match 58.7%; Score 1163; DB 1; Length 1726;
Best Local Similarity 57.5%; Pred. No. 5.3e-55;
Matches 225; Conservative 60; Mismatches 84; Indels 22; Gaps 3;

Qy 7 IDNILSKTENEYEVLYKPLAGVYRSKQLQENNVMTFNVVVKDILNSRFNFKNVL 66
Db 1338 MDNILSGFENEYDVLYKPLAGVYRSKQLQENNVMTFNVVVKDILNSRFNFKNVL 1397
Qy 67 ESDLIPYKDLTSSNVVVDYKFLAKRDKFLSYNYIKDSIDTDFINFANDVLGYKIL 126
Db 1398 ESDLQMFQKHISSEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQGGISYIEKV 1457
Qy 127 SBKYSDLSIKKYYI-----NDKQGENEKYLPFLNNIETLYKTVN 166
Db 1458 LAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNLV 1517
Qy 167 DKIDLFIHLEAKVLYNTEKSNVEVKIKELNYLTIQDKLADFKNNFVGIADLSTDY 226
Db 1518 NKIDYLYNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNHNDFDAIKKLIND 1577
Qy 227 NNHNLITFLSTGMVFENLLKSVLNSLLDWKLARYVKVHFTTPMRKKTMIQOSSGCFRHL 286
Db 1578 TKKMDLGLLSTGLV-QNPNIIISKLIEGKF-QDMLNISQHCYKQCPENSGCFRHL 1635
Qy 287 ERECKCLLYKQEGDKCVENPNTCNENNGGCDADAKTEEDSGSNKKITCECTKPCD 346
Db 1636 ERECKCLLYKQEGDKCVENPNTCNENNGGCDADAKTEEDSGSNKKITCECTKPCD 1695
Qy 347 YLFDGIFCSSNFGLISFLILMLILYSFI 377
Db 1696 YLFDGIFCSSNFGLISFLILMLILYSFI 1726

RESULT 5
MSPL_PLAF3
ID MSPL_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=8816657; PubMed=3327588;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
RL [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
RT merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
CC -|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -|- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC DE the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M35727; AAA29715.1; -
DR EMBL; Y00087; CAA68280.1; -
DR EMBL; Z35326; CAA84555.1; -
DR PIR; S06286; S06286.
DR InterPro; IPR000561; -
DR Pfam; PF00008; BGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82ALEI59948CAD6 CRC64;

Query Match 58.4%; Score 1156.5; DB 1; Length 1682;
Best Local Similarity 57.7%; Pred. No. 1.1e-54;
Matches 225; Conservative 58; Mismatches 86; Indels 21; Gaps 3;

Qy 7 IDNILSKTENEYEVLYKPLAGVYRSKQLQENNVMTFNVVVKDILNSRFNFKNVL 66
Db 1295 MDNILSGFENEYDVLYKPLAGVYRSKQLQENNVMTFNVVVKDILNSRFNFKNVL 1354
Qy 67 ESDLIPYKDLTSSNVVVDYKFLAKRDKFLSYNYIKDSIDTDFINFANDVLGYKIL 126
Db 1355 ESDLQMFQKHISSEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQGGISYIEKV 1414
Qy 127 SBKYSDLSIKKYYI-----NDKQGENEKYLPFLNNIETLYKTVD 167
Db 1415 LAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNLV 1474
Qy 168 KIDLEVIHLEAKVLYNTEKSNVEVKIKELNYLTIQDKLADFKNNFVGIADLSTDY 227
Db 1475 KIDYLYNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNPYDEAIKKLIND 1534
Qy 228 HNNLITKFLSTGMVFENLLKSVLNSLLDWKLARYVKVHFTTPMRKKTMIQOSSGCFRHL 287
Db 1535 KMDLGLLSTGLV-QNPNIIISKLIEGKF-QDMLNISQHCYKQCPENSGCFRHL 1592
Qy 288 REECKCLLYKQEGDKCVENPNTCNENNGGCDADAKTEEDSGSNKKITCECTKPCY 347
Db 1593 REECKCLLYKQEGDKCVENPNTCNENNGGCDADAKTEEDSGSNKKITCECTKPCY 1652
Qy 348 YLFDGIFCSSNFGLISFLILMLILYSFI 377
Db 1653 YLFDGIFCSSNFGLISFLILMLILYSFI 1682

RESULT 6
MSPL_PLAF3
ID MSPL_PLAF3 STANDARD; PRT; 1701 AA.
AC P13819;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA).
GN MSP-1.
```



```
FT CARBOHYD 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2BC59AF96EA98 CRC64;

Query Match 57.8%; Score 1145; DB 1; Length 1701;
Best Local Similarity 56.8%; Pred. No. 4.7e-54;
Matches 222; Conservative 61; Mismatches 86; Indels 22; Gaps 3;

QY 7 IDNILSKIEYEVLYKPLAGYRSLKQLENNVMTFNVVVKDILNSRFNKNFKNVL 66
DB 1313 MDNILSGFENEYDVLYKPLAGYRSLKQLENNVMTFNVVVKDILNSRLKRRKYELDLV 1372
QY 67 ESDLIPIKDLTSSNYVVDYKFLNKEKDKFLSSNYIKDSIDTDFINANDVLGYKIL 126
DB 1373 ESDLMQFKHSSNEYIEDSFLLNSEQKNTLLSKYIKESVENDIKFAEGISYIEKV 1432
QY 127 SEKYKSDLSIKKI-----NDKOGENEKYLPLFNNITETLYKTVN 166
DB 1433 LAKYKDDLESIKKVIKEEKEKFPSPPTTPSPAKTDEQKESKFLPFTNITETLYNNLV 1492
QY 167 DKIDLFVHLEAKVLYTYEKSNEVYKIKELNYLKTIOQKLADFKKNNNFVGIADLSTDY 226
DB 1493 NKIDDLYLINAKKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNFEAIKKLIND 1552
QY 227 NHNNLTKFLSGMFWPENLLKSVLSNLLDWKLARYKHTTTPMRKMTMQSSGGRPHLD 286
DB 1553 TRKMDLGLKLLTGLV-QNPNPTIISKLEGF-QDMLNISQHCQVKQCPKSGGRPHLD 1610
QY 287 ERECKCLNLYKQEGDKGVENPNPTCNENNGGCDADAKTEEDSGNGKKITCECTKPCD 346
DB 1611 ERECKCLNLYKQEGDKGVENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTKPCS 1670
QY 347 YPLFDGIFCSSNFLGISFLILMLILYSFI 377
DB 1671 YPLFDGIFCSSNFLGISFLILMLILYSFI 1701

RESULT 8
MSPL_PLAYO STANDARD; PRT; 1772 AA.
AC P13828;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (230 KDA).
GN MSP-1.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205979; PubMed=2320061;
RA Lewis A.P.;
RT "Sequence analysis upstream of the gene encoding the precursor to the
major merozoite surface antigens of Plasmodium yoelii."
RL Mol. Biochem. Parasitol. 39:285-288(1990).
RN [2]
RP SEQUENCE OF 1093-1772 FROM N.A.
RC STRAIN=17XL;
RX MEDLINE=88124889; PubMed=2448778;
RA Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
antigen encodes the epitope recognized by a protective monoclonal
antibody."
RL Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC -1- SURCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
```

```
CC ----- MEROZOITES, THE MATURATION TAKE PLACE DURING SCHIZONT. -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03612; AAA29762.1; -
DR EMBL; J04668; AAA29702.1; -
DR PIR; A28121; A28121.
DR PIR; A45532; A45532.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
FT SIGNAL 1 18
FT CHAIN 19 1772
FT CARBOHYD 54 54
FT CARBOHYD 406 406
FT CARBOHYD 646 646
FT CARBOHYD 829 829
FT CARBOHYD 1018 1018
FT CARBOHYD 1090 1090
FT CARBOHYD 1408 1408
FT CARBOHYD 1446 1446
FT CARBOHYD 1541 1541
FT CARBOHYD 1629 1629
FT CARBOHYD 1680 1680
FT CONFLICT 1521 1521
SQ SEQUENCE 1772 AA; 197230 MW; 9A6291658E0F45D CRC64;

Query Match 29.7%; Score 588.5; DB 1; Length 1772;
Best Local Similarity 34.4%; Pred. No. 2.4e-24;
Matches 132; Conservative 79; Mismatches 142; Indels 31; Gaps 7;

QY 9 NILSKTENEYEVLYKPLAGYRSLKQLENNVMTFNVVVKDILNSRFNKNFKNVL 68
DB 1401 DILSETNESLYVYTKRLGSTYKSLKHKMLREFSTIKEDMTNGLNNKSKRNDLFLEVL 1460
QY 69 DLIPYKDLTSSNYVVDYKFLNKEKDKFLSSNYIKDSIDTDFINANDVLGYKILSE 128
DB 1461 ELDLFKDLSTNKYVIRNPYQLLNDKDKQIVNLKYATKGINEDIETTTDGKFFNKWVE 1520
QY 129 KYKSDLSIKKI-----ND-KOGENEKYLPLFNNITETLYKTVNDKIDLFVHLEAKV 180
DB 1521 LYNTQLAAVKEQIATEAETNTNKEKKYIPILEDKGLYETVIGQAEEYSEELQNL 1580
QY 181 LNVYVEKSNVEYKIKELNYLKTIOQKLADF-----KKNNFVGIADLSTDYNNHNLTKFL 236
DB 1581 DNYKKEKAEEFILTAKNLEKYOIDEKLFVEHAENKKNHIAIA-----LNNLN 1629
QY 237 STGMVFENLLKSVLSNLLDWK----LARYVKHFTTPMRKKTMIQOSSGCFRHLDERECK 292
DB 1630 KSLVGESEKILAKMLNMDGMDLLGVDPKHVCVDTRD---IPKNAGCFRDDNGTEWR 1686
QY 293 CLUNTKQ-BGDCKVENPNPTCNENNGGCDADAKCTEEDSGSGKKITCECTKPCYPLD 351
DB 1687 CLLGKKGEGNCTVENNPNPTCDINNGGCDPTASCONAESTENSKKIICCTKPTPAYVE 1746
QY 352 GIFCSSNPLGISFLILMLILYS 375
DB 1747 GVFCSSSPFMGLSILLIITLIVFN 1770

RESULT 9
RBP2_PLAYO STANDARD; PRT; 1251 AA.
ID RBP2_PLAYO
AC Q00799;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
```


RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: BELONGS TO THE SPECTRIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U14433; AAA27973.1; -
DR WormPep; C50C3.2; CE01861.
DR InterPro; IPR002017; -
DR InterPro; IPR002048; -
DR Pfam; PF00435; Spectrin; 12.
DR PROSITE; PS00018; EF_HAND; 1.
KW Hypothetical protein; Calcium-binding.
FT CA_BIND 2025 2036 POTENTIAL.
SQ SEQUENCE 2198 AA; 256260 MW; CD1B2C1092C5EDC8 CRC64;

Query Match 6.9%; Score 136.5; DB 1; Length 2198;
Best Local Similarity 21.3%; Pred. No. 4;
Matches 69; Conservative 66; Mismatches 98; Indels 91; Gaps 15;
QY 8 DNILSKIENEYEVLYLKPLAGYRSLSKLE-----NNVTNFNVYKDIILNSRFNKR 59
DB 72 NNVLRSWEYLLKLLALK-----WRLEKESIEIPKNCDSIVDWTLTKTE---SDESK 123
QY 60 ENFNKVLSDLIPIYKDLTSSNVVVDYKFLNKRKDFLSSYNIKDSIDTIDFANDV 119
DB 124 EIFDEIRKCSMW-WNEISATFQQLDP-----TATDT 154
QY 120 LGYKILS--EYKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVHLE 177
DB 155 SNYKVKHETWSNFOEYDLSLHK-----KLSENERFQKVDNAEDLIKWMDDK----- 201
QY 178 AKVLNTYKSNVEYKIKELNYLKTIDQKLADFKNNNFVGVADLSTDYNHNNLLTKFLS 237
DB 202 EKEICEKYSKMFVEWMK---YRKVEIEMKSVQKQ-----IDLKEQF----- 242
QY 238 TGMVFENLLKSVLSNLDWK-----LARYVKHFTTPMRK-KTMIOQSSGCFRHLDEREC 291
DB 243 VGMENDN-----LRNIPDPKNHVIDVEQRFESFQAFVQKWKTDIENSADADKLMEAEANI 297
QY 292 KCLLNYKQEGDKCV-ENPNPTCNE 314
DB 298 CCWSSEKIEDLKIMATSDTPDCDE 321

RESULT 12
RPOP_AGABT
ID RPOP_AGABT STANDARD; PRT; 1102 AA.
AC P33539;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6).
OS Agaricus bitorquis.

OG Mitochondrion.
OG Plasmid pEM.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Agaricales;
OC Agaricaceae; Agaricus.
OX NCBI_TaxID=5343;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 24666 / AG4;
RX MEDLINE=91347410; PubMed=1879001;
RA Robison M.M., Royer J.C., Horgen P.A.;
RT "Homology between mitochondrial DNA of Agaricus bisporus and an
RT internal portion of a linear mitochondrial plasmid of Agaricus
RT bitorquis.";
RL Curr. Genet. 19:495-502(1991).
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -!- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X63075; CAA44799.1; -
DR PIR; S28104; S28104.
DR InterPro; IPR002092; -
DR Pfam; PF00940; RNA_pol; 1.
DR PROSITE; PS00489; RNA_POL_PHASE_2; 1.
DR PROSITE; PS00900; RNA_POL_PHASE_1; 1.
KW Transfrase; Transcription; DNA-directed RNA polymerase;
KW Mitochondrion; Plasmid.
FT ACT_SITE 734 734 BY SIMILARITY.
FT ACT_SITE 804 804 BY SIMILARITY.
FT ACT_SITE 980 980 BY SIMILARITY.
SQ SEQUENCE 1102 AA; 126664 MW; F72FDD308D1AB5B4 CRC64;

Query Match 6.8%; Score 135; DB 1; Length 1102;
Best Local Similarity 22.5%; Pred. No. 2,2;
Matches 73; Conservative 57; Mismatches 131; Indels 64; Gaps 14;
QY 5 SVIDNLSKIENEYEVLYLKPLAGYRSLSKLENNVTNFNVYKDIILNSRFNKR-NPK 63
DB 27 SLVDNLFYNLHD-----YAS---QSKHNIIDDKGKLLNLCNVLDRNVN 70
QY 64 NVLESDLPIYKDLTS--SNVYKDPYKFLNKEK---ROKFLSSYNIKD---SIDTD-IN 114
DB 71 ELNDSVFLIKDIEKECKYVSNVLYRSVGRKNGREVESYKYNKFNKVLDKGIIS 130
QY 115 FANDVLGYKILSEYKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVNDKIDLFVI 174
DB 131 INNEVLKFSKEREGERIESIAVTN-----KILENNIAEVLISIKNV----- 179
QY 175 HLEAKVLNYYTEKSNVEYKIKELNYLKTIDQKLADFKNNNF-----VGIADLSTDYNNH 229
DB 180 ----IVLN----KESVAKVEEINY--EVHNFKINGNTNFSNRNLTEIKSLKELNKM 229
QY 230 NLLTFKLTGMYFENLLKSVLSNLDWKIARYVKHFTTPMRKKTMIQSSGCFRHLDERE 289
DB 230 EILDNRINKLSTKESDLKVKIILDSNLIIDKOLAI---EKTVVEVELTFFRNMMDTH 286
QY 290 ECKCLLNYKQEGDKCVENPNPTCNE 314
DB 287 ETR-----NKIHNLYPKLNK 302

QY 142 NDKQGENEYLPFLNNIETLYKTVDKIDLVFVHLEAKVNTYKSNVEVKIKELNYLK 201
 | | | | | : : : | | | : : : | | | : : : | | | : : :
 Db 323 RSKSEFAKMEYLE-----YYQFYKELLNGNYKA-----LNFLQ 359
 QY 202 TTDKDLADFKNNFVGADLSTYDYNHNNLLTKF-----LSTG-----MVFENLLKSVLSN 252
 : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
 Db 360 KAKKSLKNVKISNIIFSRYTLECD-----ELICRFYLSISQGEFKKSVLLDEVLEISLKI 415
 QY 253 LLDWKLARYVKHF 265
 | | | | | : : : | | | : : : | | | : : : | | | : : :
 Db 416 LSDWKNTRYKXY 428

RESULT 15
 CHEA_BORBU STANDARD; PRT: 864 AA.
 AC Q44737; P70857; Q44877;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CHEMOTAXIS PROTEIN CHEA (EC 2.7.3.-).
 GN CHEA OR BB0669.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxID=139;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=212;
 RX MEDLINE=98438936; PubMed=9765799;
 RA Trueba G.A., Old I.G., Saint-Girons I., Johnson R.C.;
 RT "A cheA cheW operon in Borrelia burgdorferi, the agent of Lyme
 disease.";
 RL Res. Microbiol. 148:191-200(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Utterback T., Wattley L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi.";
 RL Nature 390:580-586(1997).
 RN [3]
 RP SEQUENCE OF 1-30 FROM N.A.
 RC STRAIN=212;
 RX MEDLINE=97144545; PubMed=8990312;
 RA Ge Y., Charon N.W.;
 RT "An unexpected flaA homolog is present and expressed in Borrelia
 burgdorferi.";
 RL J. Bacteriol. 179:552-556(1997).
 CC -!- FUNCTION: INVOLVED IN THE TRANSMISSION OF SENSORY SIGNALS FROM
 CC THE CHEMORECEPTORS TO THE FLAGELLAR MOTORS. CHEA IS
 CC AUTOPHOSPHORYLATED; IT CAN TRANSFER ITS PHOSPHATE GROUP TO EITHER
 CC CHEB OR CHEY (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
 CC KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U28962; AAB96835.1; -
 DR EMBL; AE001168; AAC67024.1; -
 DR EMBL; U62900; AAC44771.1; -
 DR EMBL; X91907; CAA63002.1; -
 DR TIGR; BB0669; -
 DR InterPro; IPR000410; -
 DR InterPro; IPR002545; -
 DR InterPro; IPR002570; -
 DR Pfam; PF01584; Chew; 1.
 DR Pfam; PF01627; Hpt; 1.
 DR Pfam; PF00512; signal; 1.
 DR PRINTS; PR00344; BCTRLSENSOR.
 KW Sensory transduction; Transferase; Kinase; Phosphorylation;
 Chemotaxis.
 FT MOD_RES 51 51 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CONFLICT 11 11 L -> I (IN REF. 1).
 FT CONFLICT 182 182 S -> H (IN REF. 1).
 FT CONFLICT 187 187 S -> G (IN REF. 1).
 FT CONFLICT 239 239 L -> S (IN REF. 1).
 FT CONFLICT 362 362 A -> S (IN REF. 1).
 FT CONFLICT 559 559 S -> P (IN REF. 1).
 SQ SEQUENCE 864 AA; 98352 MW; C1111DE0877BE624 CRC64;
 Query Match 6.7%; Score 133; DB 1; Length 864;
 Best Local Similarity 20.3%; Pred. No. 2.2;
 Matches 84; Conservative 64; Mismatches 121; Indels 144; Gaps 19;
 QY 8 DNILSKIENEYVLYLKPLAGVYRSLSKQLENNVMTFNVNVDILNSRNKKNFNKLVLE 67
 | | | | | : : : | | | : : : | | | : : : | | | : : :
 Db 214 DKFLKRV--DYVLIY--SDIEGVKSLD-----SLNL-IESYLVDEFNVEELKKLAD 261
 QY 68 SLLIPVKDL-TSSNVVVKDPYFLNKEKDKFLSSYN-----YIKSDIDYDINFAN 117
 : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
 Db 262 EEI---KQVLDLSNFVINDNFDFTEDEISDLLLEVENQKLFKVRDLFVKNPMTATIS-GL 317
 QY 118 DVLGYKILSEKYSKSDLSIKKYINDKQGENEYLPFLNNIETLYKTVDKIDL--FVH 175
 : | | | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
 Db 318 QMLQALSKLGKIFKSPIDS-SELLADKFFDVIYIISNTSE---ESIAKKINLPDVVSH 373
 QY 176 LEAKVLNYYTEKSNVEVKIKELNYLKTQDKLADFKKNNFVGADLSTD----- 225
 | | | | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
 Db 374 FEIKNVNLSLKS---VRLKE-----DDEAPFKNKNKKNKSPISVNLIRIDSKID 422
 QY 226 -----YNNH-----NLLTKFLSTGMVE 243
 : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
 Db 423 YILNLVSEAVISKSSYNQINSEMILFYNFYDYQESFORNFLIDLVKIVFDAGLTLE 482
 QY 244 NLLKSVLSNLLDWKLARYVKHFT-----TPMRK---KTMIQ 276
 : : : : : | | | : : : | | | : : : | | | : : : | | | : : :
 Db 483 DEIESHINSLSMFKMEKALKDISELRNSPFRLLQNFKMTSGRLSRIITDLHESVLKTRML 542
 QY 277 QSSGGFRHL-----DERECKCLLNYKQBGDK-----CVEN 307
 | | | | | : : : | | | : : : | | | : : : | | | : : : | | | : : :
 Db 543 PISNIFSRFTRVVRDLSSKLNKLNKLNKMEGETELDKSVIDDLVDPMLMHCVRN 595

Search completed: August 8, 2001, 12:32:42
 Job time: 210 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:58 ; Search time 43.63 seconds
(without alignments)
658.212 Million cell updates/sec

Title: US-09-500-376-5

Perfect score: 1982

Sequence: 1 AVTTSVIDNLSKIENEYEV.....SNFLGISFLLILMLILYSFI 377

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1947	98.2	1631	1 SAZQK1	major merozoite su
2	1942	98.0	1640	2 A24594	probable major sur
3	1824.5	92.1	400	2 A55545	major merozoite su
4	1824.5	92.1	1639	2 S05603	major merozoite su
5	1167.5	58.9	651	2 SA7282	merozoite surface
6	1164	58.7	1726	1 SAZQGM	major merozoite su
7	1157	58.4	1726	2 A5948	major merozoite su
8	1145	57.8	1701	2 A54498	major merozoite su
9	1131	57.1	1701	2 A26868	major merozoite su
10	745	37.6	1726	2 A39401	merozoite surface
11	744	37.5	1751	2 A45604	major blood-stage
12	588.5	29.7	680	2 A28121	major merozoite su
13	588.5	29.7	1772	2 A45532	major merozoite su
14	546	27.5	1785	2 A45546	major merozoite su
15	157.5	7.9	3394	2 T18501	hypothetical prote
16	155.5	7.8	2269	2 T28677	rhoptyr protein -
17	155.5	7.8	3973	2 B71612	hypothetical prote
18	154	7.8	2010	2 B71616	phosphatase (acid
19	153.5	7.7	1306	2 T28313	ORF MSV152 probabl
20	152.5	7.7	2401	2 T28676	rhoptyr protein -
21	149	7.5	507	2 A71622	hypothetical prote
22	149	7.5	722	2 D71607	VP45-like protein
23	149	7.5	980	2 E71606	hypothetical prote
24	149	7.5	1346	2 G71613	hypothetical prote
25	148.5	7.5	706	2 A57719	dihydropteroate sy
26	148.5	7.5	1619	2 T18499	hypothetical prote
27	147.5	7.4	1127	2 T28317	ORF MSV156 hypothe
28	145	7.3	1714	2 E71609	Ser/Thr protein ki
29	144.5	7.3	1119	2 T18491	hypothetical prote

30	142	7.2	481	2 T18465	hypothetical prote
31	142	7.2	1252	2 B42771	reticulocyte-bind
32	142	7.2	1939	2 T18372	repeat orcanellar
33	142	7.2	3844	2 T18402	asparagine/asparta
34	141	7.1	608	2 T28301	ORF MSV140 hypothe
35	141	7.1	1005	2 A64465	hypothetical prote
36	141	7.1	1187	2 T18355	hypothetical prote
37	141	7.1	2500	2 G71609	hypothetical prote
38	140.5	7.1	2708	2 T09079	probable chloroqui
39	140.5	7.1	2819	2 T09080	probable chloroqui
40	140	7.1	1308	2 E71622	probable membrane
41	139.5	7.0	377	2 H64387	hypothetical prote
42	139	7.0	2523	2 T18477	hypothetical prote
43	138.5	7.0	2013	2 G71610	probable membrane
44	138.5	7.0	2166	2 G70163	hypothetical prote
45	138	7.0	821	2 S67087	hypothetical prote

ALIGNMENTS

RESULT 1

SAZQK1

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C:Accession: A25120

R:MacKay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.
EMBO J. 4, 3823-3829, 1985

A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f
A:Reference number: A91030; MUID:86136024

A:Accession: A25120

A:Molecule type: DNA

A:Residues: 1-1631 <MAC>

C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Superfamily: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>
F:67-84/Region: 3-residue repeats (S-G-T/P)

F:1614-1631/Domain: membrane anchor #status predicted <MBN>
F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn

Query Match 98.2%; Score 1947; DB 1; Length 1631;
Best local Similarity 98.4%; Pred. No. 2.8e-95;
Matches 371; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AVTTSVIDNLSKIENEYEVLYLKLPLAGVYRSLKKOLENNVMTFNVNVDILNSRENKRE 60

DB 1255 AVTPSVIDNLSKIENEYEVLYLKLPLAGVYRSLKKOLENNVMTFNVNVDILNSRENKRE 1314

QY 61 NFKNVLESLLIPYKDLTSSNVYVDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVL 120

DB 1315 NFKNVLESLLIPYKDLTSSNVYVDPYKFLNKEKRDKFLSSYNYIKDSITDINFANDVL 1374

QY 121 GYKILSEYKSDLSIKKYINDKQGENEKLPLFLNIETLYKTVNDKIDLFIHLEAVK 180

DB 1375 GYKILSEYKSDLSIKKYINDKQGENEKLPLFLNIETLYKTVNDKIDLFIHLEAVK 1434

QY 181 LNVYTEKSNEVKIKELNLYLKTQDKLADFKNNNVFVGLADLSTDYNNHNLTKFLSTGM 240

DB 1435 LNVYTEKSNEVKIKELNLYLKTQDKLADFKNNNVFVGLADLSTDYNNHNLTKFLSTGM 1494

QY 241 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQSSGCFRHLDERECKCLLYKOE 300

DB 1495 VFENLLKSVLSNLLDWKLARYVKHFTTPMRKKTMIQSSGCFRHLDERECKCLLYKOE 1554

QY 301 GSKCVENPNTCNENNGGCDAAKCTEEDSGSGKKITCECTKPCDYPFLDFGIFCSSNF 360

DB 1555 GSKCVENPNTCNENNGGCDAAKCTEEDSGSGKKITCECTKPCDYPFLDFGIFCSSNF 1614

QY 361 LGISFLILMLILYSFI 377
 |||||
 Db 1615 LGISFLILMLILYSFI 1631
 |||||
 RESULT 2
 A24594
 probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
 C:Accession: A24594
 R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, R.; Blackman, M.J.; Ling, I.T.; Holder, A.A.
 Nature 317, 270-273, 1985
 A:Title: Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum
 A:Reference number: A24594; MUID:86014355
 A:Accession: A24594
 A:Molecule type: DNA
 A:Residues: 1-1640 <HOL>
 C:Superfamily: major merozoite surface antigen
 C:Keywords: surface antigen

Query Match 98.0%; Score 1942; DB 2; Length 1640;
 Best Local Similarity 98.7%; Pred. No. 5.2e-95;
 Matches 372; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AVTTSVIDNLSKIENEYEVLYKPLAGYRSLLKQLENNVMTFNVVVKDILNSRFNKR 60
 |||||
 Db 1264 AVTTSVIDNLSKIENEYEVLYKPLAGYRSLLKQLENNVMTFNVVVKDILNSRFNKR 1323
 |||||
 QY 61 NFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSNYIKDSITDITDINFANDVL 120
 |||||
 Db 1324 NFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSNYIKDSITDITDINFANDVL 1383
 |||||
 QY 121 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDKIDLEVIHLEAKV 180
 |||||
 Db 1384 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDKIDLEVIHLEAKV 1443
 |||||
 QY 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDYNHNNLLTKFLSTGM 240
 |||||
 Db 1444 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDYNHNNLLTKFLSTGM 1503
 |||||
 QY 241 VFENLLKSVLSNLLDWKLARYKHFTTPMRKKTMIQQSSGCFRHLDERECKCLLNKYQE 300
 |||||
 Db 1504 VFENLLKSVLSNLLDWKLARYKHFTTPMRKKTMIQQSSGCFRHLDERECKCLLNKYQE 1563
 |||||
 QY 301 GOKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPLFDGIFCSSSNF 360
 |||||
 Db 1564 GSKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPLSMVIFCSSSNF 1623
 |||||
 QY 361 LGISFLILMLILYSFI 377
 |||||
 Db 1624 LGISFLILMLILYSFI 1640
 |||||

RESULT 3
 A45545
 major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)
 C:Species: Plasmodium falciparum
 C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
 C:Accession: A45545
 R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
 Mol. Biochem. Parasitol. 49, 29-33, 1991
 A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1
 A:Reference number: A45545; MUID:92131048
 A:Accession: A45545
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-400 <BLA>
 A:Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBI:77621)
 C:Superfamily: major merozoite surface antigen
 C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 92.1%; Score 1824.5; DB 2; Length 400;
 Best Local Similarity 93.6%; Pred. No. 1.7e-89;
 Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEYEVLYKPLAGYRSLLKQLENNVMTFNVVVKDILNSRFNKR 60
 |||||
 Db 25 AVTSPVIDNLSKIENEYEVLYKPLAGYRSLLKQLENNVMTFNVVVKDILNSRFNKR 84
 |||||
 QY 61 NFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSNYIKDSITDITDINFANDVL 120
 |||||
 Db 85 NFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSNYIKDSITDITDINFANDVL 144
 |||||
 QY 121 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDKIDLEVIHLEAKV 180
 |||||
 Db 145 GYKILSEKYSKDLDSIKKYINDKOGENEKYLPLFNNIETLYKTVDKIDLEVIHLEAKV 204
 |||||
 QY 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDYNHNNLLTKFLSTGM 240
 |||||
 Db 205 LNYTEKSNVEVKIKELNYLKTIOQKLADFKKNNFVGIADLSTDYNHNNLLTKFLSTGM 264
 |||||
 QY 241 VFENLLKSVLSNLLDWKLARYKHFTTPMRKKTMIQQSSGCFRHLDERECKCLLNKYQE 300
 |||||
 Db 265 VFENLLKSVLSNLLDWKLARYKHFTTPMRKKTMIQQSSGCFRHLDERECKCLLNKYQE 323
 |||||
 QY 301 GOKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPLFDGIFCSSSNF 360
 |||||
 Db 324 GOKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCECTKPCYPLFDGIFCSSSNF 383
 |||||
 QY 361 LGISFLILMLILYSFI 377
 |||||
 Db 384 LGISFLILMLILYSFI 400
 |||||
 RESULT 4
 S05603
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
 N:Alternate names: gp195 surface antigen
 C:Species: Plasmodium falciparum
 C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
 C:Accession: S05603; S04850
 R:Myler, P.J.
 submitted to the EMBL Data Library, April 1989
 A:Reference number: S05603
 A:Accession: S05603
 A:Molecule type: mRNA
 A:Residues: 1-1639 <MYL>
 A:Cross-references: EMBL:X15063; NID:99896; PIDN:CAA33163.1; PID:99897
 R:Myler, P.J.
 Nucleic Acids Res. 17, 5401, 1989
 A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from P1
 A:Reference number: S04850; MUID:89345116
 A:Accession: S04850
 A:Molecule type: mRNA
 A:Residues: 1504-1639 <MYL2>
 A:Cross-references: EMBL:X15063
 C:Superfamily: major merozoite surface antigen
 C:Keywords: glycoprotein; merozoite; surface antigen
 F:1-19/domain: signal sequence #status predicted <SIG>
 F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 92.1%; Score 1824.5; DB 2; Length 1639;
 Best Local Similarity 93.6%; Pred. No. 8.2e-89;
 Matches 353; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEYEVLYKPLAGYRSLLKQLENNVMTFNVVVKDILNSRFNKR 60
 |||||
 Db 1264 AVTSPVIDNLSKIENEYEVLYKPLAGYRSLLKQLENNVMTFNVVVKDILNSRFNKR 1323
 |||||
 QY 61 NFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSNYIKDSITDITDINFANDVL 120
 |||||
 Db 1324 NFKNVLESLLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSNYIKDSITDITDINFANDVL 1383
 |||||

```
QY 121 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTNDKIDLFIHLEAKV 180
|||||
Db 1384 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTNDKIDLFIHLEAKV 1443
QY 181 LNYTEKSNVEVKIKELNYLKTQKDLADFKNNNFVGIADISTDYNNHNLTKFLSTGM 240
|||||
Db 1444 LNYTEKSNVEVKIKELNYLKTQKDLADFKNNNFVGIADISTDYNNHNLTKFLSTGM 1503
QY 241 VFENILKSVLSNLLDVKLARYVYKHFTTPMRKKTMTIQOSSGCFRHLDERECKCLNLYKOE 300
|||||
Db 1504 VFENILKSVLSNLLDGNL-OGMLNISQHCVKKQPQNSGCFRHLDERECKCLNLYKOE 1562
QY 301 GDKCVENPNTCNENNGGDADAKTEEDSGNGKKITCECTKPCYPLFDGIFCSSNF 360
|||||
Db 1563 GDKCVENPNTCNENNGGDADAKTEEDSGNGKKITCECTKPCYPLFDGIFCSSNF 1622
QY 361 LGISFLILMLILYSFI 377
|||||
Db 1623 LGISFLILMLILYSFI 1639

RESULT 5
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (4
C:Species: Plasmodium falciparum
A:Variety: strain RO-71
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C:Accession: S47282
R.Toller, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A:Description: Plasmodium falciparum: recombination within the C-terminal region of mer
A:Reference number: S47282
A:Accession: S47282
A:Molecule type: DNA
A:Residues: 1-651 <TOL>
A:Cross-references: EMBL:235329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A:Experimental source: strain RO-71
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 58.9%; Score 1167.5; DB 2; Length 651;
Best Local Similarity 57.9%; Pred. No. 1.3e-54;
Matches 226; Conservative 59; Mismatches 84; Indels 21; Gaps 3;

QY 7 IDNILSKTENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVYKDLINSLRFRNFKNVYL 66
|||||
Db 264 MDNILSGFENEYDVLYLPLAGVYRSLLKQLENNVMTFNVVYKDLINSLRFRNFKNVYL 323
QY 67 ESDLIPYKDLTSSNVVVDYKFLNKRDKFLSSVNYIKDSIDTDINFANDVLGYKIL 126
|||||
Db 67 ESDLIPYKDLTSSNVVVDYKFLNKRDKFLSSVNYIKDSIDTDINFANDVLGYKIL 126
QY 324 ESDLQMFHISSEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQGGISYIEKV 383
|||||
Db 324 ESDLQMFHISSEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQGGISYIEKV 383
QY 127 SEKYSDLSIKKYI-----NDKQGENEKYLPFLNNIETLYKTND 167
|||||
Db 384 LAKYKDDLESIKKVIKEKEFPSPPTTPSPAKTDEQKESKFLPFLTNIETLYNNLV 443
QY 168 KIDLFIHLEAKVLYNTEKSNVEVKIKELNYLKTQKDLADFKNNNFVGIADISTDY 227
|||||
Db 444 KIDYDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKKLINDT 503
QY 228 HNNLLTKFLSTGMVFENLLKSVLSNLLDVKLARYVYKHFTTPMRKKTMTIQOSSGCFRHLDE 287
|||||
Db 504 KIDMGLKLLSTGLV-QNPNNTIISKLEGEK-ODMLNISQHCVKKQPQNSGCFRHLDE 561
QY 288 REECKCLNLYKQEGDKCVENPNTCNENNGGDADAKTEEDSGNGKKITCECTKPCY 347
|||||
Db 562 REECKCLNLYKQEGDKCVENPNTCNENNGGDADAKTEEDSGNGKKITCECTKPCY 347
QY 348 PLFDGIFCSSNFIKISFLILMLILYSFI 377
|||||
Db 622 PLFDGIFCSSNFIKISFLILMLILYSFI 651
```

```
RESULT 6
SAZQM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: 195K glycoprotein
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361
R.Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A:Title: Variation in the gene encoding a major merozoite surface antigen of the huma
A:Reference number: A23386; MUID:86205236
A:Accession: A23386
A:Molecule type: DNA
A:Residues: 1-1104 <WEB1>
A:Cross-references: EMBL:X03831
R.Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria
A:Reference number: S06361; MUID:88143999
A:Accession: S06361
A:Molecule type: DNA
A:Residues: 1104-1726 <WEB2>
A:Cross-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1726/Product: major merozoite surface antigen #status predicted <MAT>
F:67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
F:757-765/Region: 3-residue repeats (T-E-E)
F:133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb

Query Match 58.7%; Score 1164; DB 1; Length 1726;
Best Local Similarity 57.5%; Pred. No. 5.9e-54;
Matches 225; Conservative 60; Mismatches 84; Indels 22; Gaps 3;

QY 7 IDNILSKTENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVYKDLINSLRFRNFKNVYL 66
|||||
Db 1338 MDNILSGFENEYDVLYLPLAGVYRSLLKQLENNVMTFNVVYKDLINSLRFRNFKNVYL 1397
QY 67 ESDLIPYKDLTSSNVVVDYKFLNKRDKFLSSVNYIKDSIDTDINFANDVLGYKIL 126
|||||
Db 67 ESDLIPYKDLTSSNVVVDYKFLNKRDKFLSSVNYIKDSIDTDINFANDVLGYKIL 126
QY 1398 ESDLQMFHISSEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQGGISYIEKV 1457
|||||
QY 127 SEKYSDLSIKKYI-----NDKQGENEKYLPFLNNIETLYKTND 166
|||||
Db 1458 LAKYKDDLESIKKVIKEKEFPSPPTTPSPAKTDEQKESKFLPFLTNIETLYNNLV 1517
QY 167 KIDLFIHLEAKVLYNTEKSNVEVKIKELNYLKTQKDLADFKNNNFVGIADISTDY 226
|||||
Db 1518 NKIDYDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKKLINDD 1577
QY 227 HNNLLTKFLSTGMVFENLLKSVLSNLLDVKLARYVYKHFTTPMRKKTMTIQOSSGCFRHLDE 286
|||||
Db 1578 TKKDMGLKLLSTGLV-QNPNNTIISKLEGEK-ODMLNISQHCVKKQPQNSGCFRHLDE 1635
QY 287 REECKCLNLYKQEGDKCVENPNTCNENNGGDADAKTEEDSGNGKKITCECTKPCD 346
|||||
Db 1636 REECKCLNLYKQEGDKCVENPNTCNENNGGDADAKTEEDSGNGKKITCECTKPCD 346
QY 347 YPLFDGIFCSSNFIKISFLILMLILYSFI 377
|||||
Db 1696 YPLFDGIFCSSNFIKISFLILMLILYSFI 1726

RESULT 7
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
```

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C:Accession: A45948
R:Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major merozoite surface antigen
A:Reference number: A45948; MUID:89005525
A:Accession: A45948
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1726 <CHA>
A:Cross-references: GB:M37213
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 58.4%; Score 1157; DB 2; Length 1726;
Best Local Similarity 57.3%; Pred. No. 1.4e-53;
Matches 224; Conservative 60; Mismatches 85; Indels 22; Gaps 3;

QY 7 IDNILSKIENEYEVLYLKLPLAGVYRSLKQLENNVMTFNVVVKDILNSRFRKRNFKNVL 66
DB 1338 MDNILSGFENEYDVLYLKLPLAGVYRSLKQLENNVMTFNVVVKDILNSRFRKRNFKNVL 1397
QY 67 ESDLIPYKDLTSSNVVYKDPYKFLNKEKRDKFLSSNYIKSDIDTDINPANDVLGYKIL 126
DB 1398 ESDLMOFKHISSNYEIEDSFLLNSEOKNTLLKSYKIKESVENDIKPAQEGISYK 1457
QY 127 SEKYKSLDLSIKKYI-----NDKOGENEKYLPLFNNIETLYKTVN 166
DB 1458 LAKYKDDLESIKKVIKEKEKFPSPPTTPSPVKTDEQKESKFLPFTNIETLYNNLV 1517
QY 167 DKIDLFIHLEAKVLYNYEKSNEVVKIKELNYLKTIOQKADFKKNNFVGIADLSTDY 226
DB 1518 NKDDYLYNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFRNHNDPEAIKKLIND 1577
QY 227 NNHLLTKFLSTGMVFNLLKSVLSNLDKWLARYVKKHTTPMRKKTMTQOSSGGRHL 286
DB 1578 TKKDLMLGKLLTGLV-QNFPNTIISKLEGGK-QDMLNISQHCYKOCPCNSGGRHL 1635
QY 287 ERECKCLLYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTCTCKPDC 346
DB 1636 ERECKCLLYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTCTCKPDC 1695
QY 347 YPLFDGIFCSSNFGISFLILMLILYSFI 377
DB 1696 YPLFDGIFCSSNFGISFLILMLILYSFI 1726

RESULT 8
A:Accession: A54498
C:Species: Plasmodium falciparum
C>Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.; Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmodium falciparum
A:Reference number: A54498; MUID:88142999
A:Accession: A54498
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PET>
A:Cross-references: GB:M19143; PIDN:AAA29653.1; PID:g160413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

DB 1313 MDNILSGFENEYDVLYLKLPLAGVYRSLKQLENNVMTFNVVVKDILNSRFRKRNFKNVL 1372
QY 67 ESDLIPYKDLTSSNVVYKDPYKFLNKEKRDKFLSSNYIKSDIDTDINPANDVLGYKIL 126
DB 1373 ESDLMOFKHISSNYEIEDSFLLNSEOKNTLLKSYKIKESVENDIKPAQEGISYK 1432
QY 127 SEKYKSLDLSIKKYI-----NDKOGENEKYLPLFNNIETLYKTVN 166
DB 1433 LAKYKDDLESIKKVIKEKEKFPSPPTTPSPVKTDEQKESKFLPFTNIETLYNNLV 1492
QY 167 DKIDLFIHLEAKVLYNYEKSNEVVKIKELNYLKTIOQKADFKKNNFVGIADLSTDY 226
DB 1493 NKDDYLYNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFRNHNDPEAIKKLIND 1552
QY 227 NNHLLTKFLSTGMVFNLLKSVLSNLDKWLARYVKKHTTPMRKKTMTQOSSGGRHL 286
DB 1553 TKKDLMLGKLLTGLV-QNFPNTIISKLEGGK-QDMLNISQHCYKOCPCNSGGRHL 1610
QY 287 ERECKCLLYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTCTCKPDC 346
DB 1611 ERECKCLLYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTCTCKPDC 1670
QY 347 YPLFDGIFCSSNFGISFLILMLILYSFI 377
DB 1671 YPLFDGIFCSSNFGISFLILMLILYSFI 1701

RESULT 9
A:Accession: A26868
C:Species: Plasmodium falciparum
C>Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C:Accession: A26868
R:Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum
A:Reference number: A26868; MUID:88011243
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <TAN>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 57.1%; Score 1131; DB 2; Length 1701;
Best Local Similarity 56.5%; Pred. No. 3.2e-52;
Matches 222; Conservative 60; Mismatches 85; Indels 26; Gaps 4;

QY 7 IDNILSKIENEYEVLYLKLPLAGVYRSLKQLENNVMTFNVVVKDILNSRFRKRNFKNVL 66
DB 1313 MDNILSGFENEYDVLYLKLPLAGVYRSLKQLENNVMTFNVVVKDILNSRFRKRNFKNVL 1372
QY 67 ESDLIPYKDLTSSNVVYKDPYKFLNKEKRDKFLSSNYIKSDIDTDINPANDVLGYKIL 126
DB 1373 ESDLMOFKHISSNYEIEDSFLLNSEOKNTLLKSYKIKESVENDIKPAQEGISYK 1432
QY 127 SEKYKSLDLSIKKYI-----NDKOGENEKYLPLFNNIETLYKTVN 166
DB 1433 LAKYKDDLESIKKVIKEKEKFPSPPTTPSPVKTDEQKESKFLPFTNIETLYNNLV 1492
QY 167 DKIDLFIHLEAKVLYNYEKSNEVVKIKELNYLKTIOQKADFKKNNFVGIADLSTDY 226
DB 1493 NKDDYLYNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFRNHNDPEAIKKLIND 1552
QY 227 NNHLLTKFLSTGMV-FENLLKSVLSNLDKWLARYVKKHTTPMRKKTMTQOSSGGRHL 284
DB 1553 TKKDLMLGKLLTGLVQIFN---TIISKLEGGK-QDMLNISQHCYKOCPCNSGGRHL 1608
QY 285 LDERECKCLLYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTCTCKP 344
DB 1609 LDERECKCLLYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTCTCKP 1668

A:Map position: 3
A:Note: C0760c

Query Match 7.9%; Score 157.5; DB 2; Length 3394;
Best Local Similarity 24.8%; Pred. No. 1.5;
Matches 90; Conservative 61; Mismatches 121; Indels 91; Gaps 22;

QY 8 DNILSKIENEYEVLYLKPLAGVYRSK-----KOLENNVMTFNVNVKDILNS 54
Db 1201 DKIL-RLTKKYEOL-----LEENYEKIKNKEEYHACIKELEMNLERYN-NEKIILDE 1254

QY 55 RFNKNFNKNVL-ESDLI---PYKDLTSSVYVVDYPYKFLNKEKRDKFLSSVNYIKDSID 110
Db 1255 QINEKEKINIINEKYLIILEKEEYEQNKNIQIENLEKEK--KQLQEEIIQKDMIN 1312

QY 111 TDINFAN-DVLGYKILSEKYSKSDLSI-----KKYINDK-----QGENEKYLPF---- 154
Db 1313 VKLNEKNCDIKIY----EKEQYHLTLQENKDSHNYLKDKPFENLLNLNEK-LKYDHDH 1367

QY 155 -LNNIETLY--KTVNDKIDLVIHLEAKVNLVNYEKSNEVEVKIKEL-NYLKTIQDKLAD- 209
Db 1368 SLNKNITLWLEKENNKNYF--HMN---NLRVENNNLLKMKELQNKYNIKKELNER 1421

QY 210 -----FKKNNNFVGAD-----LSTDYNNHNNLLTKFLSTGCVFENLKSIVLSNL 253
Db 1422 IKQINVFERNVSTLSLRDNRSTRGSIHQINNMVMNNTHL-GPMGASKINNINISNLYYSNM 1480

QY 254 LDWKLARYVYKHFTPMRKKTMTIOQSSGCFRHLDERECKCLLNYKQEGDKCVENPNPTCN 313
Db 1481 I-----HMSHRGSIKKNKEDAEKNS-----TQARMNNDKSDTONIINNIIHNTDN 1523

QY 314 ENN 316
Db 1524 INN 1526

Search completed: August 8, 2001, 12:31:02
Job time: 110 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:01 ; Search time 34.41 Seconds
(without alignments)
225.590 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNILSKTENEYEV.....SNFLGISFLILMLILYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2.6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2.6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2.6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2.6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2.6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296	14.9	53	1 US-08-290-919-3	Sequence 3, Appli
2	277	14.0	53	1 US-08-290-919-4	Sequence 4, Appli
3	243	12.3	106	1 US-08-290-919-11	Sequence 11, Appli
4	201	10.1	48	1 US-08-290-919-1	Sequence 1, Appli
5	201	10.1	48	1 US-08-290-919-12	Sequence 12, Appli
6	198	10.0	48	1 US-08-290-919-2	Sequence 2, Appli
7	125.5	6.3	2391	2 US-08-446-855A-2	Sequence 2, Appli
8	125.5	6.3	2391	4 US-09-150-741-2	Sequence 2, Appli
9	124	6.3	1786	4 US-08-973-462-8	Sequence 8, Appli
10	122	6.2	3135	1 US-08-323-170B-2	Sequence 2, Appli
11	121.5	6.1	984	1 US-08-257-073-3	Sequence 3, Appli
12	121.5	6.1	984	2 US-08-184-009-120	Sequence 120, App
13	121.5	6.1	984	2 US-08-458-356-120	Sequence 120, App
14	119.5	6.0	1435	2 US-08-568-459A-4	Sequence 4, Appli
15	119.5	6.0	1435	2 US-08-487-826B-4	Sequence 4, Appli
16	119.5	6.0	1494	3 US-08-755-587-186	Sequence 186, App
17	118	6.0	978	2 US-08-415-593-43	Sequence 43, Appli
18	117	5.9	455	2 US-08-392-625-21	Sequence 21, Appli
19	117	5.9	455	2 US-08-466-961A-21	Sequence 21, Appli
20	117	5.9	455	2 US-08-645-193B-23	Sequence 23, Appli
21	116.5	5.9	1147	3 US-08-470-260-5	Sequence 5, Appli
22	116.5	5.9	1147	3 US-08-471-491-5	Sequence 5, Appli
23	116.5	5.9	1147	4 US-08-466-662-5	Sequence 5, Appli
24	116.5	5.9	1147	4 US-08-477-451-2	Sequence 2, Appli
25	113.5	5.7	1239	2 US-08-937-931-2	Sequence 2, Appli
26	113.5	5.7	1239	4 US-09-285-502-2	Sequence 2, Appli
27	112.5	5.7	897	4 US-08-974-549A-189	Sequence 189, App

28	112.5	5.7	1007	3 US-08-851-843A-86	Sequence 86, Appli
29	112.5	5.7	1007	4 US-08-974-549A-187	Sequence 187, App
30	112.5	5.7	1031	3 US-08-851-843A-2	Sequence 2, Appli
31	112.5	5.7	1031	4 US-08-974-549A-110	Sequence 110, App
32	110.5	5.6	2710	1 US-08-480-604A-6	Sequence 6, Appli
33	110.5	5.6	2710	2 US-08-405-496A-6	Sequence 20, Appli
34	109.5	5.5	990	2 US-08-392-625-20	Sequence 20, Appli
35	109.5	5.5	990	2 US-08-466-961A-20	Sequence 15, Appli
36	109.5	5.5	990	2 US-08-645-193B-15	Sequence 15, Appli
37	109.5	5.5	2366	1 US-08-480-604A-10	Sequence 10, Appli
38	109.5	5.5	2366	2 US-08-405-496A-10	Sequence 10, Appli
39	108.5	5.5	903	2 US-08-209-521-24	Sequence 24, Appli
40	108.5	5.5	903	4 US-08-961-810-134	Sequence 134, App
41	108.5	5.5	903	4 US-08-352-902D-134	Sequence 134, App
42	108	5.4	439	2 US-08-853-659A-37	Sequence 37, Appli
43	107	5.4	778	5 PCT-US93-03076-4	Sequence 4, Appli
44	106.5	5.4	599	2 US-08-910-551B-2	Sequence 2, Appli
45	106.5	5.4	720	2 US-08-840-236-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MUL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site

LOCATION: 1
OTHER INFORMATION: /label= X
OTHER INFORMATION: /note= "X" = M and N, or "N"
US-08-290-919-3

Query Match 14.9%; Score 296; DB 1; Length 53;
Best Local Similarity 98.1%; Pred. No. 1.8e-16;
Matches 51; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 308 PNPTCNENGGCDADAKCTEEDSGSGNGKKTCTCTKPCDYPPLFDGIFCSSN 359
Db 2 PNPTCNENGGCDADAKCTEEDSGSGNGKKTCTCTKPCDYPPLFDGIFCSSN 53

RESULT 2

US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; APPLICATION DATA:
; PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELEPHONE: (202) 861-3000
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X" = M and N, or "N"

Query Match 14.0%; Score 277; DB 1; Length 53;
Best Local Similarity 92.3%; Pred. No. 5.3e-15;

Matches 48; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 308 PNPTCNENGGCDADAKCTEEDSGSGNGKKTCTCTKPCDYPPLFDGIFCSSN 359
Db 2 PNPTCNENGGCDADATCTEEDSGSGSRKKTCTCTKPCDYPPLFDGIFCSSN 53

RESULT 3

US-08-290-919-11
; Sequence 11, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELEPHONE: (202) 861-3000
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-290-919-11

Query Match 12.3%; Score 243; DB 1; Length 106;
Best Local Similarity 48.0%; Pred. No. 5.4e-12;
Matches 47; Conservative 14; Mismatches 33; Indels 4; Gaps 2;

QY 263 KHFTTPMRKKTMIQSSGCFRHLDEREECKLLNKKO-EGDKCVENPNPTCNENGGCD 321
Db 12 KHVCVDTRD---IPKAGCFRDNCTGTEWRCLLGYKKGEGNFCVENNPTCDINNGGCDP 68
QY 322 DAKCTEEDSGSGNGKKTCTCTKPCDYPPLFDGIFCSSN 359
Db 69 TASCQNAESTENSKKIICCTCKEPTNAVYEGVFCSSS 106

RESULT 4
US-08-290-919-1

```
; Sequence 1, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= x
; US-08-290-919-1
;
; Query Match 10.1%; Score 201; DB 1; Length 48;
; Best Local Similarity 87.5%; Pred. No. 3.3e-09;
; Matches 35; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 272 KTWIQSSGCFRHLDERECKLLNKKQEGDKCVENPNT 311
| 1:|||||||||||||||||||||||||||||||||
Db 9 KKQCPNCGCFRHLDERECKLLNKKQEGDKCVENPNT 48

RESULT 5
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= x
; US-08-290-919-1
;
; Query Match 10.1%; Score 201; DB 1; Length 48;
; Best Local Similarity 87.5%; Pred. No. 3.3e-09;
; Matches 35; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 272 KTWIQSSGCFRHLDERECKLLNKKQEGDKCVENPNT 311
| 1:|||||||||||||||||||||||||||||||||
Db 9 KKQCPNCGCFRHLDERECKLLNKKQEGDKCVENPNT 48

RESULT 5
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-290-919-12
```

```
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-290-919-12
;
; Query Match 10.1%; Score 201; DB 1; Length 48;
; Best Local Similarity 87.5%; Pred. No. 3.3e-09;
; Matches 35; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 272 KTWIQSSGCFRHLDERECKLLNKKQEGDKCVENPNT 311
| 1:|||||||||||||||||||||||||||||||||
Db 9 KKQCPNCGCFRHLDERECKLLNKKQEGDKCVENPNT 48

RESULT 6
US-08-290-919-2
; Sequence 2, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-290-919-12
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/WJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X = M and N, or N"
US-08-290-919-2

Query Match      10.0%; Score 198; DB 1; Length 48;
Best Local Similarity 85.0%; Pred. No. 5.6e-09;
Matches 34; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 272 KTWIQSSCFRHLDERECKLLNYKQBGDKCVENPNPT 311
DB 9 KQCPENSCFRHLDERECKLLNYKQBGDKCVENPNPT 48

RESULT 7
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
```

```
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-446-855A-2

Query Match      6.3%; Score 125.5; DB 2; Length 2391;
Best Local Similarity 19.3%; Pred. No. 0.37;
Matches 68; Conservative 62; Mismatches 111; Indels 111; Gaps 15;

QY 37 LENNV---MTFNVNVKDIILNSRFKNFKNVLESGLIPYKDLTSSNVV-----KD 85
DB 1333 IENNICHDISINKNIKVTINNSNISNNENV-ETNLNCVSRAGSHHIYKKEKSIGSD 1391
QY 86 PYKFLNKEKRDKFLSSYNIKDSIDTDIN-FANDVLGYIKLSEKYKSDLSIKY---- 140
DB 1392 DTNLSAQNSSNNFSCNNENMKANVDVNVLENDT-----KKREDINTTTFMEGQ 1442
QY 141 ---INDKQGENEKYLPFLNNIETLYKTVDKIDLFVHLEAKVLANVYKESNVVKIKEL 197
DB 1443 NSVINNNKKNSSLLK-----GDEEDIVMVNLK-KENNYNSVINNVDCRKKDM 1489
QY 198 NYLKTIQDKLADFKK-----NNFVGIALDSTDYHNHLLTKFLSTGMVFENLLASV 249
DB 1490 DG-KNINDECKTYKKYKMDGLNINIYDELNSGTSHTNDHL----- 1531
QY 250 LSNLLDWKLARYVKHFTTP-----MKKWTMIQSSG-----CFRHLDE 287
DB 1532 -----YLDNFNTSDEEIGNNKNMDMYLSKEKSISKNKPGNSYYVVDVSYNNEYK 1580
QY 288 REECKLLNYKQBGDKCVENPNPTC-NENNGGCDADAK-----CTEED 329
DB 1581 INKMKELIDNENLNDEYNNVNMNCSNNNASAFVNGKDRNDLENDCIEKN 1632

RESULT 8
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
US-09-150-741-2

Query Match      6.3%; Score 125.5; DB 4; Length 2391;
Best Local Similarity 19.3%; Pred. No. 0.37;
```


RESULT 11

US-08-257-073-3
; Sequence 3, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-257-073-3

Query Match 6.1%; Score 121.5; DB 1; Length 984;
Best Local Similarity 22.3%; Pred. No. 0.23;
Matches 88; Conservative 54; Mismatches 149; Indels 103; Gaps 20;

QY 7 IDNLSKIENEYEVLYLKPLA-----GVYRSLLKOLENNVMTFNVDKILNSRFNKREN 61
DB 239 LQNICETGKFKLVYIKENTLIIKVKYGETKDTENN---KVDVRKYL---INERET 291
QY 62 -PKNVLESLLIPYKDLTSSNYVVKDPYKFLN--KEKRDKFLSSYNYKIDSDTDFN--- 115
DB 292 PFTSIL---IHAYKEHNGTNLIESKNYALGSDIPEKCDT-LASNCFLSGNFNIEKCFQCA 347
QY 116 -----ANDVLGYKILSEKSDLSITKYINDKQGENEYKLPFLNNIET---LYKT 164
DB 348 LLVEKENKDVCC--YKLSIEDIVSKFEKKAETDDEDDDYETKLTESIDNILVNMFT 405
QY 165 --VNDKIDLVHLEAKVLNLYTEKSNVEVKIKELNYLKTIQD-----KLADFKNK--- 213

DB 406 NENNDKSELIKL-----EEVDDSLKLELMNYCSLLKDVDTTGTLDNYGMCNEMD 454
QY 214 ---NNFVGIIADLSTDYNNHNNLLKFLSTGMVFENLLKSVLSNLLDWKLARYVKHFTTPMRK 271
DB 455 IFNNLKRLLIYHSEENINTLKNKFRNAAV-----CLKNVDW-----IVNK 495
QY 272 KTWIQSSCCFRHLDERECKLLANYKOE-----GDKCVENPNTCNENNGGCDADAK 324
DB 496 RGLV-----LPE-----LNYDLEYFNEHLYNDR--NSPEDKDNKGGVYVHVDPT 537
QY 325 CTEEDSGSNGKKITCECTKPCYPLFDGIFCSSL 358
DB 538 LEKEDTLSYDSDNMECNKEYCNRLKDNENNCISN 571
RESULT 12
US-08-184-009-120
; Sequence 120, Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,009
; FILING DATE: 19-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-184-009-120

Query Match 6.1%; Score 121.5; DB 2; Length 984;
Best Local Similarity 22.3%; Pred. No. 0.23; Mismatches 149; Indels 103; Gaps 20;
Matches 88; Conservative 54

QY 7 IDNLSKIENEYEVLYLKPLA-----GVYRSLLKOLENNVMTFNVDKILNSRFNKREN 61
DB 239 LQNICETGKFKLVYIKENTLIIKVKYGETKDTENN---KVDVRKYL---INERET 291
QY 62 -PKNVLESLLIPYKDLTSSNYVVKDPYKFLN--KEKRDKFLSSYNYKIDSDTDFN--- 115
DB 292 PFTSIL---IHAYKEHNGTNLIESKNYALGSDIPEKCDT-LASNCFLSGNFNIEKCFQCA 347

; HYPOTHETICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium falciparum
 US-08-568-459A-4

Query Match 6.0%; Score 119.5; DB 2; Length 1435;
Best Local Similarity 21.0%; Pred. No. 0.55;
Matches 81; Conservative 46; Mismatches 123; Indels 135; Gaps 19;

[illegible]

RESULT 15

US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NH121.001CPI

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:33:44 ; Search time 70.07 Seconds
(without alignments)
708.069 Million cell updates/sec

Title: US-09-500-376-4

Perfect score: 1967

Sequence: 1 AVTPSVIHNLISKIENEYEV.....SNFLGIFILLMLILYSFI 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1767.5	89.9	539	5	Q25971 plasmodium
2	1767.5	89.9	539	5	Q25973 plasmodium
3	1763.5	89.7	400	5	Q03999 plasmodium
4	1763.5	89.7	539	5	Q25966 plasmodium
5	1763.5	89.7	539	5	Q25976 plasmodium
6	1763.5	89.7	539	5	Q25984 plasmodium
7	1763.5	89.7	539	5	Q25981 plasmodium
8	1760.5	89.5	539	5	Q25972 plasmodium
9	1760.5	89.5	539	5	Q25981 plasmodium
10	1113.5	56.6	651	5	Q25924 plasmodium
11	1110.5	56.5	569	5	Q25969 plasmodium
12	1110.5	56.5	569	5	Q25974 plasmodium
13	1110.5	56.5	569	5	Q25975 plasmodium
14	1110.5	56.5	569	5	Q25977 plasmodium
15	1110.5	56.5	569	5	Q25979 plasmodium
16	1110.5	56.5	1694	5	Q9TZT5 plasmodium
17	1110.5	56.5	1694	5	Q9NHX1 plasmodium
18	1110.5	56.5	1704	5	Q9TZT4 plasmodium
19	1109.5	56.4	569	5	Q25983 plasmodium

ALIGNMENTS

```

RESULT 1
Q25971 ID Q25971 PRELIMINARY; PRT; 539 AA.
AC Q25971;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwee S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13359; BAA02620.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001245; -.
DR Pfam; PF00008; EGF. 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61017 MW; 5CA9C651BB62B5E1 CRC64;

```

Query Match 89.9%; Score 1767.5; DB 5; Length 539;
Best Local Similarity 91.8%; Pred. No. 6.8e-89;
Matches 346; Conservative 7; Mismatches 21; Indels 3; Gaps 2;

```

QY 1 AVTPSVIHNLISKIENEYEVLYLKPLAGVYRSKQLQLENNVMTFNVVVKDLNSPFNKR 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 AVTPSVIDNLSKIENEYEVLYLKPLAGVYRSKQLQLENNVMTFNVVVKDLNSRKNR 223
QY 61 NFNKLVESDLIPYKDLTSSNVVYKDPYKFLNKKRKFLLSYNYIKSDITDINFANDVL 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 NFNKLVESDLIPYKDLTSSNVVYKDPYKFLNKKRKFLLSYNYIKSDITDINFANDVL 283

```

Q25967 plasmodium
Q25970 plasmodium
Q25980 plasmodium
Q25982 plasmodium
Q25968 plasmodium
Q25978 plasmodium
Q25922 plasmodium
Q25923 plasmodium
Q25724 plasmodium
Q25725 plasmodium
Q25721 plasmodium
Q43997 plasmodium
Q25726 plasmodium
Q43995 plasmodium
Q25722 plasmodium
Q25723 plasmodium
Q25717 plasmodium
Q25719 plasmodium
Q25720 plasmodium
Q25727 plasmodium
Q43996 plasmodium
Q25718 plasmodium
Q25728 plasmodium
Q9UB87 plasmodium
Q9NCN2 plasmodium

```
QY 121 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 180
DB 284 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 343
QY 181 LNTYKSNVE--IKELIYKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
DB 344 LNTYKSNVEVKIKELNYLTKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 403
QY 239 VFENLLKSIILNLDLWKLVYVHFTTPMRKKTMIQONSGCFRHLDERECKCLLNYKQE 298
DB 404 VFENLAKTVLSNLIDGNL-OGMLNISQHCQVKKQPCQNSGCFRHLDERECKCLLNYKQE 462
QY 299 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTTCQCTKPDSPYPLSMVIFCSSSNF 358
DB 463 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTTCCTKPDSPYPLDGFICSSSNF 522
QY 359 LGISFLILMLILYSFI 375
DB 523 LGISFLILMLILYSFI 539
RESULT 2
Q25973 PRELIMINARY; PRT; 539 AA.
AC Q25973
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13361; BAA02622.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001245; -.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
FT SEQUENCE 539 AA; 61047 MW; 105C2E02974FB398 CRC64;
SQ
Query Match 89.9%; Score 1767.5; DB 5; Length 539;
Best Local Similarity 91.8%; Pred. No. 6.8e-89;
Matches 346; Conservative 7; Mismatches 21; Indels 3; Gaps 2;
QY 1 AVTPSVTHNLISKIENEYEVLYLKPLAGVYRSKQLQENNVMTFNVVKDILNSFPNKRE 60
DB 164 AVTPSVIDNLISKIENEYEVLYLKPLAGVYRSKQLQENNVMTFNVVKDILNSRFNKRE 223
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSITDINFANDVL 120
DB 224 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSITDINFANDVL 283
QY 121 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 180
DB 284 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 343
QY 181 LNTYKSNVE--IKELIYKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
DB 344 LNTYKSNVEVKIKELNYLTKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 403
QY 239 VFENLLKSIILNLDLWKLVYVHFTTPMRKKTMIQONSGCFRHLDERECKCLLNYKQE 298
DB 404 VFENLAKTVLSNLIDGNL-OGMLNISQHCQVKKQPCQNSGCFRHLDERECKCLLNYKQE 462
QY 299 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTTCQCTKPDSPYPLSMVIFCSSSNF 358
DB 463 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTTCCTKPDSPYPLDGFICSSSNF 522
QY 359 LGISFLILMLILYSFI 375
DB 523 LGISFLILMLILYSFI 539
RESULT 3
Q03999 PRELIMINARY; PRT; 400 AA.
AC Q03999
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE MEROZOITE SURFACE ANTIGEN PRECURSOR 1 (PMMSA) (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blackman M.J., Ling I.T., Nicholls S.C., Holder A.A.;
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: COULD BE ATTACHED TO THE MEMBRANE BY A GPI-
CC ANCHOR.
CC -1- MISCELLANEOUS: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF
CC 83K, 42K & 19K ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
DR EMBL; M54681; AAA29709.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001245; -.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT NON_TER 1
FT SIGNAL <1 24 POTENTIAL.
FT CHAIN 25 400 83K MEROZOITE SURFACE ANTIGEN.
FT CHAIN 25 286 42K MEROZOITE SURFACE ANTIGEN.
FT CHAIN 287 400 19K MEROZOITE SURFACE ANTIGEN.
FT TRANSMEM 383 400 MEMBRANE ANCHOR.
FT SEQUENCE 400 AA; 45824 MW; 537F075058626AC2 CRC64;
SQ
Query Match 89.7%; Score 1763.5; DB 5; Length 400;
Best Local Similarity 91.5%; Pred. No. 8.1e-89;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;
QY 1 AVTPSVTHNLISKIENEYEVLYLKPLAGVYRSKQLQENNVMTFNVVKDILNSFPNKRE 60
DB 25 AVTPSVIDNLISKIENEYEVLYLKPLAGVYRSKQLQENNVMTFNVVKDILNSRFNKRE 84
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSITDINFANDVL 120
DB 85 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSITDINFANDVL 144
QY 121 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 180
DB 145 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 204
QY 181 LNTYKSNVE--IKELIYKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
DB 205 LNTYKSNVEVKIKELNYLTKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 264
QY 239 VFENLLKSIILNLDLWKLVYVHFTTPMRKKTMIQONSGCFRHLDERECKCLLNYKQE 298
DB 265 VFENLAKTVLSNLIDGNL-OGMLNISQHCQVKKQPCQNSGCFRHLDERECKCLLNYKQE 323
QY 299 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTTCQCTKPDSPYPLSMVIFCSSSNF 358
```

```
QY 121 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 180
DB 284 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 343
QY 181 LNTYKSNVE--IKELIYKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
DB 344 LNTYKSNVEVKIKELNYLTKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 403
QY 239 VFENLLKSIILNLDLWKLVYVHFTTPMRKKTMIQONSGCFRHLDERECKCLLNYKQE 298
DB 404 VFENLAKTVLSNLIDGNL-OGMLNISQHCQVKKQPCQNSGCFRHLDERECKCLLNYKQE 462
QY 299 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTTCQCTKPDSPYPLSMVIFCSSSNF 358
DB 463 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKTTCCTKPDSPYPLDGFICSSSNF 522
QY 359 LGISFLILMLILYSFI 375
DB 523 LGISFLILMLILYSFI 539
RESULT 2
Q25973 PRELIMINARY; PRT; 539 AA.
AC Q25973
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13361; BAA02622.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001245; -.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
FT SEQUENCE 539 AA; 61047 MW; 105C2E02974FB398 CRC64;
SQ
Query Match 89.9%; Score 1767.5; DB 5; Length 539;
Best Local Similarity 91.8%; Pred. No. 6.8e-89;
Matches 346; Conservative 7; Mismatches 21; Indels 3; Gaps 2;
QY 1 AVTPSVTHNLISKIENEYEVLYLKPLAGVYRSKQLQENNVMTFNVVKDILNSFPNKRE 60
DB 164 AVTPSVIDNLISKIENEYEVLYLKPLAGVYRSKQLQENNVMTFNVVKDILNSRFNKRE 223
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSITDINFANDVL 120
DB 224 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSSYNIKDSITDINFANDVL 283
QY 121 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 180
DB 284 GYKILSEKSDLSIKKYINDKQGENEYLPPLNNIETLYKTVNDKIDLVHLEAKV 343
QY 181 LNTYKSNVE--IKELIYKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
DB 344 LNTYKSNVEVKIKELNYLTKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 403
QY 239 VFENLLKSIILNLDLWKLVYVHFTTPMRKKTMIQONSGCFRHLDERECKCLLNYKQE 298
```

```

Db 324 GDKVFNPTCNENGGCDADAKTEEDSGNGKKTCTCKPDSPYPLFDGIFCSSNF 383
Qy 359 LGISFLILMLILYSFI 375
Db 384 LGISFLILMLILYSFI 400

RESULT 4
Q25966
ID Q25966 PRELIMINARY; PRT; 539 AA.
AC Q25966;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13357; BAA02618.1; -
DR InterPro; IPR000561; -
DR InterPro; IPR001245; -
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61144 MW; 7B7DE90C1D0ACDE7 CRC64;

Query Match 89.7%; Score 1763.5; DB 5; Length 539;
Best Local Similarity 91.5%; Pred. No. 1.1e-88;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

Qy 1 AVTPSVIHNLISKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVVKDILNSPFNKRE 60
Db 164 AVTPSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVVKDILNSRFNKRE 223
Qy 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKRKDRKFLSSYNYIKDSIDTDINFANDVL 120
Db 224 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKRKDRKFLSSYNYIKDSIDTDINFANDVL 283
Qy 121 GYKILSEKYSDDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 180
Db 284 GYKILSEKYSDDLSDISKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 343
Qy 181 LNYTEKSNVE--IKELYLKTIOQKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGM 238
Db 344 LNYTEKSNVEVKIKELNYLKTIOQKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGM 403
Qy 239 VFENLLKSILSNLNDWLKARYVKHFTTPMRKKTMTQNSGCFRHLDERECKCLLNYKQE 298
Db 404 VFENLAKTVLSNLLDGNL-QGMLNISQHCYKQCPQNSGCFRHLDERECKCLLNYKQE 462
Qy 299 GSKCVENPNTCNENGGCDADAKCTEEDSGNGKKTCTCKPDSPYPLSVIFCSSNF 358
Db 463 GDKVFNPTCNENGGCDADAKCTEEDSGNGKKTCTCKPDSPYPLFDGIFCSSNF 522
Qy 359 LGISFLILMLILYSFI 375
Db 523 LGISFLILMLILYSFI 539

RESULT 6
Q25984
ID Q25984 PRELIMINARY; PRT; 539 AA.
AC Q25984;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

```

```

RN  SEQUENCE FROM N.A.
RP  MEDLINE-93295445; PubMed=8515786;
RA  Jongwutives S., Tanabe K., Kanbara H.;
RT  "Sequence conservation in the C-terminal part of the precursor to the
RT  major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT  field isolates.";
RL  Mol. Biochem. Parasitol. 59:95-100(1993).
DR  EMBL; D13356; BAA02617.1; -.
DR  InterPro; IPR001245; -.
DR  InterPro; IPR000561; -.
DR  Pfam; PF00008; EGF; 1.
DR  PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW  Merozoite; EGF-like domain.
FT  NON_TER 1
SQ  SEQUENCE 539 AA; 61114 MW; 3788015F3127CB9E CRC64;

Query Match      89.7%; Score 1763.5; DB 5; Length 539;
Best Local Similarity 91.5%; Pred. No. 1.1e-88;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

QY  1 AVTPSVIHNLISKIENEYEVLYLKPLAGVYRSLSKKQLENNVMTFNVVKDILNSRPNKRE 60
Db  164 AVTPSVIDNLSKIENEYEVLYLKPLAGVYRSLSKKQLENNVMTFNVVKDILNSRPNKRE 223
QY  61 NFNKLVESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNIKDSITDINFANDVL 120
Db  224 NFNKLVESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNIKDSITDINFANDVL 283
QY  121 GYKILSEKYSKDSLSIKKYINDKQGENEYLPPLNNIETLYKTVDNKIDLFVHLEAKV 180
Db  284 GYKILSEKYSKDSLSIKKYINDKQGENEYLPPLNNIETLYKTVDNKIDLFVHLEAKV 343
QY  181 LNYTYEKSNEV--IKELIYLTQIDKLADFKKNNNFVGIADLSTDYNNHLLTKFLSTGM 238
Db  344 LNYTYEKSNEVVKIKELNYLTQIDKLADFKKNNNFVGIADLSTDYNNHLLTKFLSTGM 403
QY  239 VFENLKSILSNLDLWKLVYKFTTPMRKKTMIQONSGCFRHLDERECKCLLNKQK 298
Db  404 VFENLAKTVLSNLDGNL--QGLNISQHCVKQCPONSGCFRHLDERECKCLLNKQK 462
QY  299 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKITCCTKPDSPYPLSMVIFCSSNF 358
Db  463 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKITCCTKPDSPYPLSMVIFCSSNF 522
QY  359 LGISFLILMLILYSFI 375
Db  523 LGISFLILMLILYSFI 539

RESULT 7
Q9TYG1 PRELIMINARY; PRT; 539 AA.
AC Q9TYG1
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13358; BAA02619.1; -.
DR InterPro; IPR000561; -.

```

```

DR  InterPro; IPR001245; -.
DR  Pfam; PF00008; EGF; 1.
DR  PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW  Merozoite; EGF-like domain.
FT  NON_TER 1
SQ  SEQUENCE 539 AA; 61075 MW; C70C2E100EC4A101 CRC64;

Query Match      89.7%; Score 1763.5; DB 5; Length 539;
Best Local Similarity 91.5%; Pred. No. 1.1e-88;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

QY  1 AVTPSVIHNLISKIENEYEVLYLKPLAGVYRSLSKKQLENNVMTFNVVKDILNSRPNKRE 60
Db  164 AVTPSVIDNLSKIENEYEVLYLKPLAGVYRSLSKKQLENNVMTFNVVKDILNSRPNKRE 223
QY  61 NFNKLVESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNIKDSITDINFANDVL 120
Db  224 NFNKLVESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNIKDSITDINFANDVL 283
QY  121 GYKILSEKYSKDSLSIKKYINDKQGENEYLPPLNNIETLYKTVDNKIDLFVHLEAKV 180
Db  284 GYKILSEKYSKDSLSIKKYINDKQGENEYLPPLNNIETLYKTVDNKIDLFVHLEAKV 343
QY  181 LNYTYEKSNEV--IKELIYLTQIDKLADFKKNNNFVGIADLSTDYNNHLLTKFLSTGM 238
Db  344 LNYTYEKSNEVVKIKELNYLTQIDKLADFKKNNNFVGIADLSTDYNNHLLTKFLSTGM 403
QY  239 VFENLKSILSNLDLWKLVYKFTTPMRKKTMIQONSGCFRHLDERECKCLLNKQK 298
Db  404 VFENLAKTVLSNLDGNL--QGLNISQHCVKQCPONSGCFRHLDERECKCLLNKQK 462
QY  299 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKITCCTKPDSPYPLSMVIFCSSNF 358
Db  463 GSKCVENPNTCNENGGCDADAKCTEEDSGSNGKKITCCTKPDSPYPLSMVIFCSSNF 522
QY  359 LGISFLILMLILYSFI 375
Db  523 LGISFLILMLILYSFI 539

RESULT 8
Q25972 PRELIMINARY; PRT; 539 AA.
AC Q25972
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13360; BAA02621.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001245; -.
DR Pfam; PF00008; EGF; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ  SEQUENCE 539 AA; 61076 MW; 72186F1412C57D1C CRC64;

Query Match      89.5%; Score 1760.5; DB 5; Length 539;
Best Local Similarity 91.2%; Pred. No. 1.6e-88;

```



```
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 56.5%; Score 1110.5; DB 5; Length 569;
Best Local Similarity 56.7%; Pred. No. 3.7e-53;
Matches 220; Conservative 60; Mismatches 85; Indels 23; Gaps 4;

QY 9 NILSKIENEYEVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSPFNKRNFKNVLES 68
DB 184 NILSGFENEYDVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSRLKRRYFLDVLES 243
QY 69 DLIYPKDLTSSNVYVVDYKFLNKEKROKFLSSYNYKDSIDTDINFANDVLYGYKILSE 128
DB 244 DLMQFKHSSNEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAQEGISYVEKVL 303
QY 129 KYKSLDSTTKYI-----NDKOGENEKYLPLFNNIETLYKTVNDKI 169
DB 304 KYKDDLESIKKVIKEKEPPSPPTPPSPAKTDSQKESKELPFLNLIETLYNNLVNKI 363
QY 170 DLFVHLEAKVLYTYEK--SNVEIKELIYTKTIDQKLADFKNNNFVGIADLSTDYHNH 227
DB 364 DDYLNLRKAKINDCNVDEAHVKITKLSDLKAIIDKIDLFKNTNDFEAIKKLINDDTKK 423
QY 228 NLLTFSLTGWVFNLLSKLSILNLDWKLYRVKFTTPMRKKTMIQNSGCGFRHLDERE 287
DB 424 DMLGKLLSTGLV-QNPNTIISKLEGRF-QDMLNISQHCQVKKQCPENSGCGFRHLDERE 481
QY 288 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 347
DB 482 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 541
QY 348 SMVIFCSSNFGISFLILMLILYSFI 375
DB 542 FDGIFCSSNFGISFLILMLILYSFI 569

RESULT 14.
Q25977
ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 56.5%; Score 1110.5; DB 5; Length 569;
Best Local Similarity 56.7%; Pred. No. 3.7e-53;
Matches 220; Conservative 60; Mismatches 85; Indels 23; Gaps 4;

QY 9 NILSKIENEYEVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSPFNKRNFKNVLES 68
DB 184 NILSGFENEYDVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSRLKRRYFLDVLES 243
QY 69 DLIYPKDLTSSNVYVVDYKFLNKEKROKFLSSYNYKDSIDTDINFANDVLYGYKILSE 128
DB 244 DLMQFKHSSNEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAQEGISYVEKVL 303
QY 129 KYKSLDSTTKYI-----NDKOGENEKYLPLFNNIETLYKTVNDKI 169
DB 304 KYKDDLESIKKVIKEKEPPSPPTPPSPAKTDSQKESKELPFLNLIETLYNNLVNKI 363
QY 170 DLFVHLEAKVLYTYEK--SNVEIKELIYTKTIDQKLADFKNNNFVGIADLSTDYHNH 227
DB 364 DDYLNLRKAKINDCNVDEAHVKITKLSDLKAIIDKIDLFKNTNDFEAIKKLINDDTKK 423
QY 228 NLLTFSLTGWVFNLLSKLSILNLDWKLYRVKFTTPMRKKTMIQNSGCGFRHLDERE 287
DB 424 DMLGKLLSTGLV-QNPNTIISKLEGRF-QDMLNISQHCQVKKQCPENSGCGFRHLDERE 481
QY 288 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 347
DB 482 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 541
QY 348 SMVIFCSSNFGISFLILMLILYSFI 375
DB 542 FDGIFCSSNFGISFLILMLILYSFI 569

RESULT 15.
Q25979
ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13352; BAA02613.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 56.5%; Score 1110.5; DB 5; Length 569;
Best Local Similarity 56.7%; Pred. No. 3.7e-53;
Matches 220; Conservative 60; Mismatches 85; Indels 23; Gaps 4;
```

```
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 56.5%; Score 1110.5; DB 5; Length 569;
Best Local Similarity 56.7%; Pred. No. 3.7e-53;
Matches 220; Conservative 60; Mismatches 85; Indels 23; Gaps 4;

QY 9 NILSKIENEYEVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSPFNKRNFKNVLES 68
DB 184 NILSGFENEYDVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSRLKRRYFLDVLES 243
QY 69 DLIYPKDLTSSNVYVVDYKFLNKEKROKFLSSYNYKDSIDTDINFANDVLYGYKILSE 128
DB 244 DLMQFKHSSNEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAQEGISYVEKVL 303
QY 129 KYKSLDSTTKYI-----NDKOGENEKYLPLFNNIETLYKTVNDKI 169
DB 304 KYKDDLESIKKVIKEKEPPSPPTPPSPAKTDSQKESKELPFLNLIETLYNNLVNKI 363
QY 170 DLFVHLEAKVLYTYEK--SNVEIKELIYTKTIDQKLADFKNNNFVGIADLSTDYHNH 227
DB 364 DDYLNLRKAKINDCNVDEAHVKITKLSDLKAIIDKIDLFKNTNDFEAIKKLINDDTKK 423
QY 228 NLLTFSLTGWVFNLLSKLSILNLDWKLYRVKFTTPMRKKTMIQNSGCGFRHLDERE 287
DB 424 DMLGKLLSTGLV-QNPNTIISKLEGRF-QDMLNISQHCQVKKQCPENSGCGFRHLDERE 481
QY 288 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 347
DB 482 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 541
QY 348 SMVIFCSSNFGISFLILMLILYSFI 375
DB 542 FDGIFCSSNFGISFLILMLILYSFI 569

RESULT 14.
Q25977
ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 56.5%; Score 1110.5; DB 5; Length 569;
Best Local Similarity 56.7%; Pred. No. 3.7e-53;
Matches 220; Conservative 60; Mismatches 85; Indels 23; Gaps 4;

QY 9 NILSKIENEYEVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSPFNKRNFKNVLES 68
DB 184 NILSGFENEYDVLYLKPLAGVYRSLSKOLENNVMTFNVNVDILNSRLKRRYFLDVLES 243
QY 69 DLIYPKDLTSSNVYVVDYKFLNKEKROKFLSSYNYKDSIDTDINFANDVLYGYKILSE 128
DB 244 DLMQFKHSSNEYIIEDSFKLLNSEQNTLLSKYKIKESVENDIKFAQEGISYVEKVL 303
QY 129 KYKSLDSTTKYI-----NDKOGENEKYLPLFNNIETLYKTVNDKI 169
DB 304 KYKDDLESIKKVIKEKEPPSPPTPPSPAKTDSQKESKELPFLNLIETLYNNLVNKI 363
QY 170 DLFVHLEAKVLYTYEK--SNVEIKELIYTKTIDQKLADFKNNNFVGIADLSTDYHNH 227
DB 364 DDYLNLRKAKINDCNVDEAHVKITKLSDLKAIIDKIDLFKNTNDFEAIKKLINDDTKK 423
QY 228 NLLTFSLTGWVFNLLSKLSILNLDWKLYRVKFTTPMRKKTMIQNSGCGFRHLDERE 287
DB 424 DMLGKLLSTGLV-QNPNTIISKLEGRF-QDMLNISQHCQVKKQCPENSGCGFRHLDERE 481
QY 288 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 347
DB 482 ECKCLLNKQEGSKCVENPPTCNENNGCDADAKTEDSGSNGKKTCTCKPDSPYPL 541
QY 348 SMVIFCSSNFGISFLILMLILYSFI 375
DB 542 FDGIFCSSNFGISFLILMLILYSFI 569

RESULT 15.
Q25979
ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13352; BAA02613.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 56.5%; Score 1110.5; DB 5; Length 569;
Best Local Similarity 56.7%; Pred. No. 3.7e-53;
Matches 220; Conservative 60; Mismatches 85; Indels 23; Gaps 4;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:32:34 ; Search time 24.79 Seconds
(without alignments)
518.185 Million cell updates/sec

Title: us-09-500-376-4

Perfect score: 1967

Sequence: 1 AVTPSVIHNLKIEYEV.....SNFLGIFLLMLILYSFI 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1763.5	89.7	1630	1 MSP1_PLAFK	P04932 plasmodium
2	1763.5	89.7	1639	1 MSP1_PLAFW	P04933 plasmodium
3	1110	56.4	1726	1 MSP1_PLAFC	P04934 plasmodium
4	1109	56.4	1726	1 MSP1_PLAFP	P50495 plasmodium
5	1102.5	56.0	1682	1 MSP1_PLAF3	P19598 plasmodium
6	1091	55.5	1701	1 MSP1_PLAFF	P13819 plasmodium
7	1091	55.5	1701	1 MSP1_PLAFM	P08569 plasmodium
8	565.5	28.7	1772	1 MSP1_PLAYO	P13828 plasmodium
9	149.5	7.6	1162	1 BXEN_CLOBO	P46082 clostridium
10	146.5	7.4	1162	1 BXEN_CLOBU	Q06366 clostridium
11	137.5	7.0	1956	1 ATX1_PLAFA	Q04956 plasmodium
12	136.5	6.9	377	1 Y704_METJA	Q58115 methanococc
13	135	6.9	1150	1 IRR1_YEAST	P40541 saccharomyc
14	134	6.8	971	1 Y228_BORBU	O51246 borrelia bu
15	133	6.8	864	1 CHEA_BORBU	Q44737 borrelia bu
16	132	6.7	1030	1 Y018_MYCPN	P75093 mycoplasma
17	129.5	6.6	2198	1 YLJ2_CAEEL	P34367 caenorhabdi
18	129	6.6	989	1 SERA_PLAFG	P13823 plasmodium
19	129	6.6	1251	1 RBP2_PLAVB	Q00799 plasmodium
20	127.5	6.5	867	1 DP01_RICRJA	Q05949 rickettsia
21	127	6.5	756	1 Y328_MYCGE	Q49419 mycoplasma
22	126	6.4	1169	1 EX5B_BORBU	O51578 borrelia bu
23	125	6.4	442	1 TIG_BUCAI	P57546 buchnera ap
24	125	6.4	655	1 YKDA_MYCCA	P45615 mycoplasma
25	125	6.4	748	1 Y875_METJA	Q58285 methanococc
26	125	6.4	1196	1 BXCN_CLOBO	P46081 clostridium
27	124.5	6.3	749	1 MAD1_YEAST	P40957 saccharomyc
28	124.5	6.3	896	1 RPOP_NEUCR	P33540 neurospora
29	124	6.3	3305	1 APLP_MANSE	Q25490 manduca sex
30	123	6.3	881	1 LHS1_YEAST	P36016 saccharomyc
31	123	6.3	944	1 NUF1_YEAST	P32380 saccharomyc
32	122	6.2	922	1 LCN2_LACLA	P37609 lactococcus
33	122	6.2	1038	1 CIN6_YEAST	P27895 saccharomyc

RESULT 1
MSP1_PLAFK
ID MSP1_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS) (PMMSA) (P190).
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID:5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MCKLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03371; CAA27070.1; -;
DR PIR; A25120; SAZQK1.
DR InterPro; IPR000561; -;
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1630
FT DOMAIN 67 84
FT TRANSMEM 1614 1630
FT CARBOHYD 97 97
FT CARBOHYD 259 259
FT CARBOHYD 755 755
FT CARBOHYD 759 759
FT CARBOHYD 774 774
FT CARBOHYD 835 835
FT CARBOHYD 911 911
FT POTENTIAL.
FT MEROZOITE SURFACE PROTEIN 1.
FT TRIPEPTIDE SG(TP) REPEAT.
FT MEMBRANE ANCHOR.
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).

34 120.5 6.1 629 1 PR39_YEAST
35 120 6.1 1901 1 YCF1_TOBAC
36 119.5 6.1 396 1 RMAR_YEAST
37 119.5 6.1 1875 1 MLP1_YEAST
38 118.5 6.0 540 1 YK26_YEAST
39 118.5 6.0 1024 1 RPOB_PLAFA
40 118.5 6.0 1077 1 CTR9_YEAST
41 118.5 6.0 2009 1 SEC7_YEAST
42 118 6.0 703 1 HS83_PHANI
43 118 6.0 1738 1 YCF1_EPIVI
44 117.5 6.0 569 1 CYSP_PLAFA
45 117.5 6.0 766 1 YS83_CAEEL

ALIGNMENTS

P39682 saccharomyc
P12222 nicotiana t
Q02381 saccharomyc
Q02455 saccharomyc
P36112 saccharomyc
P21421 plasmodium
P89105 saccharomyc
P1075 saccharomyc
P51819 pharbitis n
Q00383 epifagus vi
P25805 plasmodium
Q09622 caenorhabdi

```
FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;

Query Match 89.7%; Score 1763.5; DB 1; Length 1630;
Best Local Similarity 91.5%; Pred. No. 8e-89;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

Qy 1 AVTPSVIHNLSKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSPFNKRE 60
Db 1255 AVTPSVIHNLSKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSPFNKRE 1314

Qy 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDRFLSYNYIKDSIDTDINFANDVL 120
Db 1315 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDRFLSYNYIKDSIDTDINFANDVL 1374

Qy 121 GYKILSEKYSKDSLSIKKYINDKQGENEYLPFLNNIETLYKTVDNKKIDLFVHLEAKV 180
Db 1375 GYKILSEKYSKDSLSIKKYINDKQGENEYLPFLNNIETLYKTVDNKKIDLFVHLEAKV 1434

Qy 181 LNTYKESNVE--IKELIYLTQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
Db 1435 LNTYKESNVEVKIKELNYLTQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 1494

Qy 239 VFENLLKSILSNLNDWKLARYKHFTTPMRKKTMIQNSGCFRHLDERECKCLLNYKQE 298
Db 1495 VFENLLKSILSNLNDGKLN--QGMLNISQHCVKQKQPCQNSGCFRHLDERECKCLLNYKQE 1553

Qy 299 GSKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTKCTCQCTKPDSPSLVWIFCSSNF 358
Db 1554 GSKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTKCTCQCTKPDSPSLVWIFCSSNF 1613

Qy 359 LGISFLILMLILYSFI 375
Db 1614 LGISFLILMLILYSFI 1630

RESULT 2
MSPI_PLAFW
ID MSPI_PLAFW STANDARD; PRT; 1639 AA.
AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A.; Lockyer M.J.; Odink K.G.; Sandhu J.S.; Riveros-Moreno V.,
RA Nicholson S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";
RL Nature 317:270-273(1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02919; CAA26676.1; -.
DR PIR; A24594; A24594.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 768 768 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C25B6616C87F6E CRC64;

Query Match 89.7%; Score 1763.5; DB 1; Length 1639;
Best Local Similarity 91.5%; Pred. No. 8e-89;
Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

Qy 1 AVTPSVIHNLSKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSPFNKRE 60
Db 1264 AVTPSVIHNLSKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSPFNKRE 1323

Qy 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDRFLSYNYIKDSIDTDINFANDVL 120
Db 1324 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKKRDRFLSYNYIKDSIDTDINFANDVL 1383

Qy 121 GYKILSEKYSKDSLSIKKYINDKQGENEYLPFLNNIETLYKTVDNKKIDLFVHLEAKV 180
Db 1384 GYKILSEKYSKDSLSIKKYINDKQGENEYLPFLNNIETLYKTVDNKKIDLFVHLEAKV 1443

Qy 181 LNTYKESNVE--IKELIYLTQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
Db 1444 LNTYKESNVEVKIKELNYLTQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 1503

Qy 239 VFENLLKSILSNLNDWKLARYKHFTTPMRKKTMIQNSGCFRHLDERECKCLLNYKQE 298
Db 1504 VFENLLKSILSNLNDGKLN--QGMLNISQHCVKQKQPCQNSGCFRHLDERECKCLLNYKQE 1562

Qy 299 GSKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTKCTCQCTKPDSPSLVWIFCSSNF 358
Db 1563 GSKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTKCTCQCTKPDSPSLVWIFCSSNF 1622

Qy 359 LGISFLILMLILYSFI 375
Db 1623 LGISFLILMLILYSFI 1639

RESULT 3
MSPI_PLAFW
ID MSPI_PLAFW STANDARD; PRT; 1726 AA.
AC P04934;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
```

DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 GN MSP-1.
 OS Plasmodium falciparum (isolate Camp / Malaysia).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5835;
 RN [1]
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE=86205236; PubMed=3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT "Variation in the gene encoding a major merozoite surface antigen of
 the human malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 14:3311-3323(1986).
 RN [2]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=88143999; PubMed=3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X03831; CAA27446.1; -;
 DR PIR: A23386; SAZOGM.
 DR InterPro: IPR000561; -;
 DR Pfam: PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196197 MW; D08AD45FA352BCF3 CRC64;
 Query Match 56.4%; Score 1110; DB 1; Length 1726;
 Best Local Similarity 56.6%; Pred. No. 2.9e-53;
 Matches 220; Conservative 60; Mismatches 85; Indels 24; Gaps 4;
 QY 9 NILSKIENEYEVLYKPLAGVYRSLLKOLENNVMTENVNKKDLNLSFPNKNFKNVLES 68
 DB 1340 NILSGFENEVDVYKPLAGVYRSLLKQIEKNIFTFNLNLDLNSLKKRYEFLDVL 1399
 QY 69 DLIPYKOLTSNNVYKDPYFLNKRDKRDLSSNYIKDSITDITDINFANDVLYKILSE 128
 DB 1400 DLMQFKHSSNEYIIEFSKLLNSEQRNLLKSKYKIKESVENDIKFAQEGISYERKLA 1459
 QY 129 KYKSDLSIKKIYI-----NDKOGENEKYLPLFNLTETLYKTVNDK 168
 DB 1460 KYKDDLESIRKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTETLYNLDVNK 1519

QY 169 IDLFVHLEAKVLNVTYK--SNVEIKELIYKTIQKLDKADFKNNNFVGIADLSTDYNH 226
 DB 1520 IDDLINLAKKINDCNVEKDEAHVKITKLSLDKAIIDKIDLFKNHDFEAIKKLINDTK 1579
 QY 227 NNLTATKFLTGMPVENLKLKLSNLLDMWKLARYVYKHTTTPMRKKTMIQOQSGCFRHLDER 286
 DB 1580 KDMGLKLLSTGLV-QNFPNTIISKLGKF-QDMLNLSQHCVKKQCPENSGCFRHLDER 1637
 QY 287 EECKLLNYKQEGSKCVENPNPTCNENNGCDADAKTEEDSGNGKKTICQCTKPDSP 346
 DB 1638 EECKLLNYKQEGDKCVENPNPTCNENNGCDADAKTEEDSGNGKKTICQCTKPDSP 1697
 QY 347 LSMVIFCSSNFGISFLILMLILYFSFI 375
 DB 1698 LFDGIFCSSNFGISFLILMLILYFSFI 1726
 RESULT 4
 MSP1_PLAPP ID MSP1_PLAPP STANDARD; PRT; 1736 AA.
 AC P50495;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (GP195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydropathy profile of the
 major merozoite surface antigen (gp195) of the Uganda-Palo Alto
 isolate.";
 RT isolate.";
 RL Exp. Parasitol. 67:1-11(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M37213; AAA29611.1; -;
 DR InterPro: IPR000561; -;
 DR Pfam: PF00008; EGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEFA2F9A026 CRC64;

Query Match 56.48; Score 1109; DB 1; Length 1726;
 Best Local Similarity 56.6%; Pred. No. 3.3e-53;
 Matches 220; Conservative 60; Mismatches 85; Indels 24; Gaps 4;

QY 9 NILSKIENEYEVLYKPLAGVYRSLLKQLENNVMTFNVNVKDIILNSPKNFKNVLES 68
 DB 1340 NILSGFENEYDVLYKPLAGVYRSLLKQLENNVMTFNVNVKDIILNSPKNFKNVLES 1399
 QY 69 DLIPYKDLTSSNVVYKDFKLNKREKDKFLSSYNYIKDSITDIDINFANDVLYGKILSE 128
 DB 1400 DLMOFKHISSEYIIEISFKLLNSEQNTLLSKYIKESVENDIKFAQEGISYKVLVLA 1459
 QY 129 KYKSDLSIKYI-----NDKQGENEKYLPFLNNIETLYKTVDNK 168
 DB 1460 KYKDDLESIKKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLNITLYNNLVNK 1519
 QY 169 IDLFVHLEAKVLNVTYK--SNVEIKELIYKTIQDLADFKKNNFVGTADLSTDYNH 226
 DB 1520 IDLVILINKAKINCENVEKDAHVYKYLKSLDKAIDDKIDFKNNHDFDAIKKLINDTK 1579
 QY 227 NLLTKFLSTGMVFENLLKSLNSLLDWKLARYVYKHTTPMRKKTMIQNSGCFRHLDER 286
 DB 1580 KDLGKLLSTGLV-QNPPNTIISKIEGK-QDMLNISQHVKKQCPENSGCFRHLDER 1637
 QY 287 ECKCLLYNKQEGSKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCQTKPDSPY 346
 DB 1638 ECKCLLYNKQEGSKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCQTKPDSPY 1697
 QY 347 LSWIFGSSSFLGISFLILMLILYSFI 375
 DB 1698 LFDGIFCSSSFLGISFLILMLILYSFI 1726

RESULT 5
 MSPI_PLAF3
 ID MSPI_PLAF3 STANDARD; PRT; 1682 AA.
 AC P19598; 025921;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate ro-33 / Ghana).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=5834;
 RN [1]
 RP SEQUENCE OF 1-1061 FROM N.A.
 RX MEDLINE=88166657; PubMed=3327688;
 RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
 RT "A naturally occurring gene encoding the major surface antigen
 RT precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
 RL EMBO J. 6:4137-4142(1987).
 RN [2]
 RP SEQUENCE OF 1032-1682 FROM N.A.
 RX MEDLINE=95354793; PubMed=7628566;
 RA Tolle R., Bujard H., Cooper J.A.;
 RT "Plasmodium falciparum: variations within the C-terminal region of
 RT merozoite surface antigen-1.";
 RL Exp. Parasitol. 81:47-54(1995).
 RN [3]
 RP SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M35727; AAA29715.1; -;
 DR EMBL; Y00087; CAA68280.1; -;
 DR EMBL; Z35326; CAA84555.1; -;
 DR PIR; S06286; S06286;
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1;
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1682
 FT TRANSMEM 1666 1682
 FT CARBOHYD 233 233
 FT CARBOHYD 462 462
 FT CARBOHYD 528 528
 FT CARBOHYD 599 599
 FT CARBOHYD 785 785
 FT CARBOHYD 881 881
 FT CARBOHYD 901 901
 FT CARBOHYD 947 947
 FT CARBOHYD 1071 1071
 FT CARBOHYD 1178 1178
 FT CARBOHYD 1569 1569
 SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;

Query Match 56.0%; Score 1102.5; DB 1; Length 1682;
 Best Local Similarity 56.7%; Pred. No. 7.2e-53;
 Matches 220; Conservative 58; Mismatches 87; Indels 23; Gaps 4;

QY 9 NILSKIENEYEVLYKPLAGVYRSLLKQLENNVMTFNVNVKDIILNSPKNFKNVLES 68
 DB 1297 NILSGFENEYDVLYKPLAGVYRSLLKQLENNVMTFNVNVKDIILNSPKNFKNVLES 1356
 QY 69 DLIPYKDLTSSNVVYKDFKLNKREKDKFLSSYNYIKDSITDIDINFANDVLYGKILSE 128
 DB 1357 DLMOFKHISSEYIIEISFKLLNSEQNTLLSKYIKESVENDIKFAQEGISYKVLVLA 1416
 QY 129 KYKSDLSIKYI-----NDKQGENEKYLPFLNNIETLYKTVDNKI 169
 DB 1417 KYKDDLESIKKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLNITLYNNLVNKI 1476
 QY 170 IDLFVHLEAKVLNVTYK--SNVEIKELIYKTIQDLADFKKNNFVGTADLSTDYNH 227
 DB 1477 IDLVILINKAKINCENVEKDAHVYKYLKSLDKAIDDKIDLVKPNYDFEAIKKLINDTKK 1536
 QY 228 NLLTKFLSTGMVFENLLKSLNSLLDWKLARYVYKHTTPMRKKTMIQNSGCFRHLDER 287
 DB 1537 DMLGKLLSTGLV-QNPPNTIISKIEGK-QDMLNISQHVKKQCPENSGCFRHLDER 1594
 QY 288 ECKCLLYNKQEGSKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCQTKPDSPY 347
 DB 1595 ECKCLLYNKQEGSKCVENPNTCNENNGGCDADAKTEEDSGNGKKITCQTKPDSPY 1654
 QY 348 SWIFGSSSFLGISFLILMLILYSFI 375
 DB 1655 FDGIFCSSSFLGISFLILMLILYSFI 1682

RESULT 6
 MSPI_PLAF3 STANDARD; PRT; 1701 AA.
 ID MSPI_PLAF3
 AC P13819;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA).
 GN MSP-1.

Plasmodium falciparum (isolate FC27 / Papua New Guinea).
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5837;
 (1)
 SEQUENCE FROM N.A.
 MEDLINE=98142999; PubMed=2449612;
 Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
 Brown G.V., Anders R.F., Kemp D.J.;
 "Variation in the precursor to the major merozoite surface antigens
 of Plasmodium falciparum";
 Mol. Biochem. Parasitol. 27:291-302(1988).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M19143; AAA29653.1; -;
 DR PIR; A54498; A54498.
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;
 Query Match 55.5%; Score 1091; DB 1; Length 1701;
 Best Local Similarity 55.8%; Pred. No. 3.le-52;
 Matches 217; Conservative 61; Mismatches 87; Indels 24; Gaps 4;
 Qy 9 NILSKIENEYEVLYKPLAGVYRSKKGLENNVMTFNVNKKDILNPFNKRNFKNVLES 68
 Db 1315 NILSGFENEYDVIYKPLAGVYRSKKGLENNVMTFNVNKKDILNPFNKRNFKNVLES 1374
 Qy 69 DLIPYKDLTSSVYVVDYKFLNKEKRDKFLSSYNYIKDSITDINANDVILGYKILSE 128
 Db 1375 DLMQFKHSSNEYIIEFSKLLNSEQNTLLSKYKIKESYENDIKFAQEGISYEVKLA 1434
 Qy 129 KYGSDLSIKKIYI-----NDKQSENEKYLPLFLNLIETLYKTVDNK 168
 Db 1435 KYKDDLESIKKVIKEEKEKFPSPPTTPPSPAKTDQOKESKFLPLFLNLIETLYNVLNK 1494
 Qy 169 IDLFVHLEAKVLYNYEK--SNVEIKELIYKTIQDKLADFKNNNFVGIADLSTDYNH 226
 Db 1495 IDDYILNLKAKINDCNVEKDEAHVKITKLSDLKIDDKIDLFKNTNDFAEAKKLINDTK 1554
 Qy 227 NNLITKFLSTGVFENLLSKLSLNLDDKLYRYKHTTPMRKTKMTQOQSGCFRHLDER 286
 Db 1555 KMLGKLLSTGLV-QNFPNTIISKLEIGKF-QDMLNISOHOCVKQCPENSGCFRHLDER 1612

Qy 287 BECKLLNLYKQEGSKCVENPNPTCNENNGGCDADAKTEEDSGNGKKITCQCTKPDSP 346
 Db 1613 BECKLLNLYKQEGSKCVENPNPTCNENNGGCDADATCTEEDSGSSKKITCCTKPDSP 1672
 Qy 347 LSWIFCSSSNFLGISFLIILMLILYSFI 375
 Db 1673 LFDGIFCSSSNFLGISFLIILMLILYSFI 1701
 RESULT 7
 MSPL_PLAFM
 ID MSPL_PLAFM STANDARD; PRT; 1701 AA.
 AC P08569;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=70153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88011243; PubMed=3079521;
 RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
 RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
 Plasmodium falciparum";
 RL J. Mol. Biol. 195:273-287(1987).
 RN [2]
 RP REVISIONS TO 1403; 1569 AND 1629.
 RA Tanabe K.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-115 FROM N.A.
 RX MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 Stunnenberg H., Bujard H.;
 RT "Polymorphism of the precursor for the major surface antigens of
 Plasmodium falciparum merozoites: studies at the genetic level.";
 RL EMBO J. 4:3823-3829(1985).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X05624; CAA29112.1; -;
 DR PIR; A26868; A26868.
 DR PIR; B25120; B25120.
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19
 FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).

```

FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 55.5%; Score 1091; DB 1; Length 1701;
Best Local Similarity 55.8%; Pred. No. 3.1e-52;
Matches 21; Conservative 61; Mismatches 87; Indels 24; Gaps 4;

Qy 9 NILSKIENEYEVLYKPLAGYVRSKQLENNVMTFNVNVDILNSPKNREKNVLES 68
Db 1315 NILSGFENEYDVLYKPLAGYVRSKQLENNVMTFNVNVDILNSPKNREKNVLES 1374

Qy 69 DLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSNYIKDSITDINFANDVGLYKILSE 128
Db 1375 DLMOFKHISSEYIETDSFKLLNSEQKNTLKSYYIKESVENDIKFAQEGISYYEKVLA 1434

Qy 129 KYKSDLSIKKYI-----NDKOGENEKYLPLFNNTIETLYKTNDK 168
Db 1435 KYKODLESIKKYIEKEKFPSSPTTPPSAKTDEQKESKFLPFTNIETLYNNLVNK 1494

Qy 169 IDLEVIHLEAKVLYTEK--SNVEIKELIYKTIQKLDKFKNNFVGIADLSTDYNH 226
Db 1495 IDLYLNLKAKINDCNVKEDEAHVKITKLSDLKAIIDKIDLFKNTDFEAIKKLINDTK 1554

Qy 227 NNLTKFLSTGVFENLLKSLNLLDVKLARYVHTTPMRKMTMQNSGCFRHLDER 286
Db 1555 KDMGLKLLSTGLV-QNFPNTIISLIEGKF-QDMLNISQHOVCVKQCPNSGCFRHLDER 1612

Qy 287 BECKLLNYKQEGKCVENPNTCNENNGGDADAKTEEDSGNGKKITCOCTKPSY 346
Db 1613 BECKLLNYKQEGKCVENPNTCNENNGGDADATCTEEDSGSRKKITCECKPSY 1672

Qy 347 LSWYIFCSSNFGISFLILMLILYSFI 375
Db 1673 LFDGIFCSSNFGISFLILMLILYSFI 1701

RESULT 8
MSPI_PLAYO STANDARD; PRT; 1772 AA.
AC P13828;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (230 KDA).
GN MSP-1.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205979; PubMed=2320061;
RA Lewis A.P.;
RT "Sequence analysis upstream of the gene encoding the precursor to the
RT major merozoite surface antigens of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 39:285-288(1990).
RN [2]
RC SEQUENCE OF 1093-1772 FROM N.A.
RP STRAIN=17XL;
RX MEDLINE=88124889; PubMed=2448778;
RA Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
RT antibody encodes the epitope recognized by a protective monoclonal
RT antibody.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

```

```

CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03612; AAA29762.1; -
DR EMBL; J04668; AAA29702.1; -
DR PIR; A28121; A28121.
DR PIR; A45532; A45532.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT CHAIN 1 18 POTENTIAL.
FT CARBOHYD 19 1772 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 829 829 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1018 1018 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1090 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1408 1408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1446 1446 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1541 1541 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1629 1629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1680 1680 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 1521 1521 L -> V (IN REF. 2).
SQ SEQUENCE 1772 AA; 197230 MW; 9A6291658EB0F45D CRC64;

Query Match 28.7%; Score 565.5; DB 1; Length 1772;
Best Local Similarity 34.9%; Pred. No. 1.2e-23;
Matches 134; Conservative 73; Mismatches 144; Indels 33; Gaps 8;

Qy 9 NILSKIENEYEVLYKPLAGYVRSKQLENNVMTFNVNVDILNSPKNREKNVLES 68
Db 1401 DLSEFTNESLYVYTKRLSGTYKSLKKHMLREFSTIKEDMTGLNKKSKRNDLEVLSH 1460

Qy 69 DLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSNYIKDSITDINFANDVGLYKILSE 128
Db 1461 ELDLFKDLSTNKYVIRNPYQLDNDKKDKQIVNLKYATKGINEDIETTDGICKFENKME 1520

Qy 129 KYKSDLSIKKYI-----ND-KOGENEKYLPLFNNTIETLYKTNDKIDLFVHLEAKV 180
Db 1521 LNTQLAAVKEQIATIEATNDTNKEKKYIPILEDKGLYETVIGQAEYSSELOLNL 1580

Qy 181 LNYTYEKSNIVEI--KELIYKTIQKLDKF---KKNNFVGIADLSTDYNHNNLTKFL 234
Db 1581 DNYKNEKAEPFILTAKLEKIQIDKLDKDFEVEHAENKHTASIA-----LNNLN 1629

Qy 235 STGMVFENLLKSLNSLLDMK----LARYVKHFTTPMRKKTMTIQNSGCFRHLDERECK 290
Db 1630 KSLGVGEGESKKILAKMLNMDGMLLVDPKHVCVDRD---IPKNAGCFRDRDNGTEWR 1686

Qy 291 CLLNYKQ-EGSKCVENPNTCNENNGGDADAKTEEDSGNSGKKITCOCTKPSYPLSM 349
Db 1687 CLLGYKKGEGNTCVENPNTCNENNGGDPTASQNAESTENSKKIITCKEPTPNAYYE 1746

Qy 350 VIFCSSNFGISFLILMLILYS 373
Db 1747 GVFCSSSFMGLSILLITLIVFN 1770

RESULT 9
BXEN_CLOBO STANDARD; PRT; 1162 AA.
ID BXEN_CLOBO
AC P46082;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)

```



```
Db 362 AKKINLPDWSHFEEK-----NNLESLSVRLKEDDEAPFKENKNIKKNSPISVN 412
Qy 224 -----YNHN-----NLTT 231
Db 413 LIRDSKKIDYILNLVSEAVISKSYNOINSEMITLTFYFNFYDYQESFORNELIDIKI 472
Qy 232 KFLSTGMVFENLLKSILSNLLDWKLARYVK 261
Db 473 VFKDAGLTLEDEIESHINSLSMFKMERALK 502
```

Search completed: August 8, 2001, 12:32:38
Job time: 206 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:54 ; Search time 43.63 Seconds
(without alignments)
654.721 Million cell updates/sec

Title: US-09-500-376-4

Perfect score: 1967

Sequence: 1 AVTPSVIHNLISKIENEYEV.....SNFLGISFLLILMLILYSFI 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1914	97.3	1631	1 SAZQK1	major merozoite su
2	1901	96.6	1640	2 A24594	probable major sur
3	1763.5	89.7	400	2 A45545	major merozoite su
4	1763.5	89.7	1639	2 S05603	major merozoite su
5	1113.5	56.6	651	2 S47282	merozoite surface
6	1110	56.4	1726	1 SAZQGM	major merozoite su
7	1103	56.1	1726	2 A45948	major merozoite su
8	1091	55.5	1701	2 A5498	major merozoite su
9	1077	54.8	1701	2 A26868	major merozoite su
10	714	36.3	1726	2 A39401	merozoite surface
11	713	36.2	1751	2 A45604	major blood-stage
12	565.5	28.7	680	2 A28121	major merozoite su
13	565.5	28.7	1772	2 A45532	major merozoite su
14	529	26.9	1785	2 A45546	major merozoite su
15	165.5	8.4	1306	2 T28313	ORF MSV152 probabl
16	164.5	8.4	2269	2 T28677	riophry protein -
17	157	8.0	1127	2 T28317	ORF MSV156 hypothe
18	156	7.9	1714	2 E71609	Ser/Thr protein ki
19	153	7.8	1346	2 G71613	hypothetical prote
20	150.5	7.7	1162	2 A47708	progenitor toxin n
21	149.5	7.6	706	2 A57119	dihydropteroate sy
22	148	7.5	507	2 A71622	hypothetical prote
23	146.5	7.4	1162	2 D140817	botulinum toxin no
24	146	7.4	722	2 D71607	VPS45-like protein
25	145.5	7.4	2401	2 T28676	riophry protein -
26	145	7.4	960	2 S72284	DNA-directed RNA p
27	145	7.4	980	2 E71606	hypothetical prote
28	143	7.3	608	2 T28301	ORF MSV140 hypothe
29	142.5	7.2	2010	2 B71616	phosphatase (acid

ALIGNMENTS

RESULT 1

SAZQK1

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C:Accession: A25120

R:Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.

EMBO J. 4, 3823-3829, 1985

A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f

A:Reference number: A91030; MUID:86136024

A:Accession: A25120

A:Molecule type: DNA

A:Residues: 1-1631 <MAC>

C:Comment: The merozoite stages of different strains have strain-specific surface ant

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>

F:67-84/Region: 3-residue repeats (S-G-T/P)

F:1614-1631/Domain: membrane anchor #status predicted <MBN>

F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn

probable chloroqui
probable chloroqui
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
probable membrane
hypothetical prote
type I restriction
p-type caton tran
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote

Query Match 97.3% Score 1914; DB 1; Length 1631;
Best Local Similarity 97.9%; Pred. No. 9.6e-97;
Matches 369; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Qy 1 AVTPSVIHNLISKIENEYEVLYKPLAGYRSKQLKLENNVMTFNVVYKDLINSPFNKRE 60

Db 1255 AVTPSVIDNLISKIENEYEVLYKPLAGYRSKQLKLENNVMTFNVVYKDLINSPFNKRE 1314

Qy 61 NFKNVLESLLPYKDLTSSVNVVYKDPYKFLNKRKDFLSSYNIKSIDIDTFNFDVL 120

Db 1315 NFKNVLESLLPYKDLTSSVNVVYKDPYKFLNKRKDFLSSYNIKSIDIDTFNFDVL 1374

Qy 121 GYKILSEKYSKSLDLSIKKYINDKQGENEYKLPFLNNTETLYKTVDNDKIDLFVHLEAKV 180

Db 1375 GYKILSEKYSKSLDLSIKKYINDKQGENEYKLPFLNNTETLYKTVDNDKIDLFVHLEAKV 1434

Qy 181 LNYTYKSNV--IKELIYKTIQDKLADFKKNNFVGIADLSTDYNNHNLTLKFLSTGM 238

Db 1435 LNYTYKSNV--IKELIYKTIQDKLADFKKNNFVGIADLSTDYNNHNLTLKFLSTGM 1494

Qy 239 VFENLLKSILNLDWLKLYRVKFTTPMRKKTMTIQQSGCFRHLDERECKLLNKKQE 298

Db 1495 VFENLLKSILNLDWLKLYRVKFTTPMRKKTMTIQQSGCFRHLDERECKLLNKKQE 1554

Qy 299 GSKCVENPNPCNENNGCCDADAKTEDSGSKKTKTQCTKPDSPSLSVIFCSSNF 358

Db 1555 GSKCVENPNPCNENNGCCDADAKTEDSGSKKTKTQCTKPDSPSLSVIFCSSNF 1614

QY 359 LGISFLILMLILYSFI 375
 |||||
 Db 1615 LGISFLILMLILYSFI 1631

RESULT 2

A24594
 probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
 C:Accession: A24594
 R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, Nature 317, 270-273, 1985
 A:Title: Primary structure of the precursor to the three major surface antigens of Plasmodium falciparum
 A:Reference number: A24594; MUID:86014355
 A:Accession: A24594
 A:Molecule type: DNA
 A:Residues: 1-1640 <HOL>
 C:Superfamily: major merozoite surface antigen
 C:Keywords: surface antigen

Query Match 96.6%; Score 1901; DB 2; Length 1640;
 Best Local Similarity 97.1%; Pred. No. 4.9e-96;
 Matches 366; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 1 AVTPSVIHNLISKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSRPNKRE 60
 |||||
 Db 1264 AVTTSVIDNLISKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSRPNKRE 1323
 |||||
 QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNYIKDSIDTDINFANDVL 120
 |||||
 Db 1324 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNYIKDSIDTDINFANDVL 1383
 |||||
 QY 121 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVJHLEAKV 180
 |||||
 Db 1384 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVJHLEAKV 1443
 |||||
 QY 181 LNTYKESNVE--IKELIYLKTIQDKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
 |||||
 Db 1444 LNTYKESNVEVKIKELNYLKTIOQKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 1503
 |||||
 QY 239 VFENLLKSILSNLNDWLKARYVHFTTPMRKKTKIQNSGCFRHLDERECKCLLNTKQE 298
 |||||
 Db 1504 VFENLLKSILSNLNDWLKARYVHFTTPMRKKTKIQNSGCFRHLDERECKCLLNTKQE 1563
 |||||
 QY 299 GSKCVENPNTCNENGGCDADAKCTEEDSGSGKKTTCQCTKPDSPYPLSMVIFCSSSNF 358
 |||||
 Db 1564 GSKCVENPNTCNENGGCDADAKCTEEDSGSGKKTTCQCTKPDSPYPLSMVIFCSSSNF 1623
 |||||
 QY 359 LGISFLILMLILYSFI 375
 |||||
 Db 1624 LGISFLILMLILYSFI 1640

RESULT 3

A45545
 major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)
 C:Species: Plasmodium falciparum
 C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
 C:Accession: A45545
 R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
 Mol. Biochem. Parasitol. 49, 29-33, 1991
 A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1
 A:Reference number: A45545; MUID:92131048
 A:Accession: A45545
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-400 <BLA>
 A:Note: sequence extracted from NCBI backbone (NCBI:77612, NCBI:77621)
 C:Superfamily: major merozoite surface antigen
 C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 89.7%; Score 1763.5; DB 2; Length 400;
 Best Local Similarity 91.5%; Pred. No. 3e-89;
 Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

QY 1 AVTPSVIHNLISKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSRPNKRE 60
 |||||
 Db 25 AVTPSVIHNLISKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSRPNKRE 84
 |||||
 QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNYIKDSIDTDINFANDVL 120
 |||||
 Db 85 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNYIKDSIDTDINFANDVL 144
 |||||
 QY 121 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVJHLEAKV 180
 |||||
 Db 145 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVJHLEAKV 204
 |||||
 QY 181 LNTYKESNVE--IKELIYLKTIQDKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
 |||||
 Db 205 LNTYKESNVEVKIKELNYLKTIOQKLADPKKNNFVGIADLSTDYNNHLLTKFLSTGM 264
 |||||
 QY 239 VFENLLKSILSNLNDWLKARYVHFTTPMRKKTKIQNSGCFRHLDERECKCLLNTKQE 298
 |||||
 Db 265 VFENLLKSILSNLNDWLKARYVHFTTPMRKKTKIQNSGCFRHLDERECKCLLNTKQE 323
 |||||
 QY 299 GSKCVENPNTCNENGGCDADAKCTEEDSGSGKKTTCQCTKPDSPYPLSMVIFCSSSNF 358
 |||||
 Db 324 GSKCVENPNTCNENGGCDADAKCTEEDSGSGKKTTCQCTKPDSPYPLSMVIFCSSSNF 383
 |||||
 QY 359 LGISFLILMLILYSFI 375
 |||||
 Db 384 LGISFLILMLILYSFI 400

RESULT 4

S05603
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
 N:Alternate names: gp195 surface antigen
 C:Species: Plasmodium falciparum
 C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
 C:Accession: S05603; S04850
 R:Myler, P.J.
 submitted to the EMBL Data Library, April 1989
 A:Reference number: S05603
 A:Accession: S05603
 A:Molecule type: mRNA
 A:Residues: 1-1639 <MYL>
 A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
 R:Myler, P.J.
 Nucleic Acids Res. 17, 5401, 1989
 A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Pl
 A:Reference number: S04850; MUID:89345116
 A:Accession: S04850
 A:Molecule type: mRNA
 A:Residues: 1504-1639 <MYL2>
 A:Cross-references: EMBL:X15063
 C:Superfamily: major merozoite surface antigen
 C:Keywords: glycoprotein; merozoite; surface antigen
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 89.7%; Score 1763.5; DB 2; Length 1639;
 Best Local Similarity 91.5%; Pred. No. 1.5e-88;
 Matches 345; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

QY 1 AVTPSVIHNLISKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSRPNKRE 60
 |||||
 Db 1264 AVTPSVIHNLISKIENEYEVLYLPLAGVYRSKQLKQLENNVMTFNVVVKDILNSRPNKRE 1323
 |||||
 QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNYIKDSIDTDINFANDVL 120
 |||||
 Db 1324 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKRDKFLSYNYIKDSIDTDINFANDVL 1383
 |||||

QY 121 GYKILSEKYSKLDLSSKKYKINDKOGENEKYLPLNNIETLYKTVDKIDLVHLEAKV 180
Db 1384 GYKILSEKYSKLDLSSKKYKINDKOGENEKYLPLNNIETLYKTVDKIDLVHLEAKV 1443
QY 181 LNYTYEKSNEV--TKELIYKTIQDKLADPKKNNNFVGIADLSTDYNNHLLTKFLSTGM 238
Db 1444 LNYTYEKSNEVKIKELNYLTKTIQDKLADPKKNNNFVGIADLSTDYNNHLLTKFLSTGM 1503
QY 239 VFENLLKSIILNLDLWKLVYVHFTTPMRKKTMTIQNSGCFRHLDERECKLLNKKOE 298
Db 1504 VFENLAKTVLSNLDGWL--QDMLNISQHCVKQCPQNSGCFRHLDERECKLLNKKOE 1562
QY 299 GSKCVENPNTCNENNGGCDADAKTEEDSGSGKKTICQCTKPDSPYPLSMVIFCSSSNF 358
Db 1563 GSKCVENPNTCNENNGGCDADAKTEEDSGSGKKTICBCTKPDSPYPLDGFICSSSNF 1622
QY 359 LGISFLLILMLILYSFI 375
Db 1623 LGISFLLILMLILYSFI 1639
RESULT 5
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (4
C:Species: Plasmodium falciparum
A:Variety: strain RO-71
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C:Accession: S47282
R:Toile, R.; Bujard, H.; Cooper, J.A.
submitted to: EMBL data library, July 1994
A:Description: Plasmodium falciparum: recombination within the C-terminal region of mer
A:Reference number: S47282
A:Accession: S47282
A:Molecule type: DNA
A:Residues: 1-651 <TOL>
A:Cross-references: EMBL:235329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A:Experimental source: strain RO-71
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
Query Match 56.6%; Score 1113.5; DB 2; Length 651;
Best Local Similarity 57.0%; Pred. No. 1.3e-53;
Matches 221; Conservative 59; Mismatches 85; Indels 23; Gaps 4;
QY 9 NILSKINEYEVLYLKLPLAGVYSLKQLENNVMTFNVNVDILNSPFNKNFKNVLES 68
Db 266 NILSGFENEYDVLYLKLPLAGVYSLKQLENNVMTFNVNVDILNSPFNKNFKNVLES 325
QY 69 DLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSVNYTKDSITDINFANDVLYGYYKTLSE 128
Db 326 DLVFIHLEAKVLYTYEK--SNVEIKELIYKTIQDKLADPKKNNNFVGIADLSTDYNNH 385
QY 129 KYKSDLSIKKYYI-----NDKOGENEKYLPLNNIETLYKTVDKID 169
Db 386 KYKDDLESIKKVIKEEKEFPSPPTTPSPAKTDEQKESKELPFLTNLETLYNNLVNKI 445
QY 170 DLVFIHLEAKVLYTYEK--SNVEIKELIYKTIQDKLADPKKNNNFVGIADLSTDYNNH 227
Db 446 DVLINLAKINDCNVEKDEAHVKITKLSLKAIDDKIDLFKNHNDPEAKKLINDDTK 505
QY 228 NLLTKFLSTGMVFENLLKSIILNLDLWKLVYVHFTTPMRKKTMTIQNSGCFRHLDERE 287
Db 506 DMLGKLLSTGLV--QNPNTIISKLEIEGKF--QDMLNISQHCVKQCPQNSGCFRHLDERE 563
QY 288 ECKCLLYNQEGSKVCVENPNTCNENNGGCDADAKTEEDSGSGKKTICQCTKPDSPYPL 347
Db 564 ECKCLLYNQEGSKVCVENPNTCNENNGGCDADAKTEEDSGSGKKTICBCTKPDSPYPL 623
QY 348 SMVIFCSSSNFNLGIFSLILMLILYSFI 375
Db 624 FDGIFCSSSNFNLGIFSLILMLILYSFI 651

RESULT 6
SAQQGM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: 195K glycoprotein
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361
R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A:Title: Variation in the gene encoding a major merozoite surface antigen of the huma
A:Reference number: A23386; MUID:86205236
A:Accession: A23386
A:Molecule type: DNA
A:Residues: 1-1104 <WEB1>
A:Cross-references: EMBL:X03831
R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria
A:Reference number: S06361; MUID:88143999
A:Accession: S06361
A:Molecule type: DNA
A:Residues: 1104-1726 <WEB2>
A:Cross-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1726/Product: major merozoite surface antigen repeats <MAT>
F:67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
F:757-765/Region: 3-residue repeats (T-E-E)
F:133,272,503,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb
Query Match 56.4%; Score 1110; DB 1; Length 1726;
Best Local Similarity 56.6%; Pred. No. 6.1e-53;
Matches 220; Conservative 60; Mismatches 85; Indels 24; Gaps 4;
QY 9 NILSKINEYEVLYLKLPLAGVYSLKQLENNVMTFNVNVDILNSPFNKNFKNVLES 68
Db 1340 NILSGFENEYDVLYLKLPLAGVYSLKQLENNVMTFNVNVDILNSPFNKNFKNVLES 1399
QY 69 DLIPYKDLTSSNVVVDYKFLNKEKRDKFLSSVNYTKDSITDINFANDVLYGYYKTLSE 128
Db 1400 DLMQFKHISNEVYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGISYIEKVL 1459
QY 129 KYKSDLSIKKYYI-----NDKOGENEKYLPLNNIETLYKTVDKID 168
Db 1460 KYKDDLESIKKVIKEEKEFPSPPTTPSPAKTDEQKESKELPFLTNLETLYNNLVNK 1519
QY 169 DLVFIHLEAKVLYTYEK--SNVEIKELIYKTIQDKLADPKKNNNFVGIADLSTDYNNH 226
Db 1520 IDLYLNLKAKINDCNVEKDEAHVKITKLSLKAIDDKIDLFKNHNDPEAKKLINDDTK 1579
QY 227 NLLTKFLSTGMVFENLLKSIILNLDLWKLVYVHFTTPMRKKTMTIQNSGCFRHLDER 286
Db 1580 KDMGLKLLSTGLV--QNPNTIISKLEIEGKF--QDMLNISQHCVKQCPQNSGCFRHLDER 1637
QY 287 ECKCLLYNQEGSKVCVENPNTCNENNGGCDADAKTEEDSGSGKKTICQCTKPDSPY 346
Db 1638 ECKCLLYNQEGSKVCVENPNTCNENNGGCDADAKTEEDSGSGKKTICBCTKPDSPY 1697
QY 347 LSWIFCSSSNFNLGIFSLILMLILYSFI 375
Db 1698 LFDGIFCSSSNFNLGIFSLILMLILYSFI 1726
RESULT 7
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum

```

Db      1315 NILSGFENEVDVYLKPLAGVYRSKKQIEKNIITFNLNLDILNSLRKKRYFDLVLES 1374
QY      69 DLIPYKDLTSSNYVVKPYKFLNKEKDKLSSYNYIKDSIDTDINFANDVLGYKYLSE 128
Db      1375 DLMOFKHISNEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGISYIEKVL 1434
QY      129 KYKSDLSISIKYI-----NDKOGENEKYLFPFLNNIETLYKTVNDK 168
Db      1435 KYKDDLESIKKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNLVNK 1494
QY      169 IDLPIVHLEAKVLNVTYEK--SNVEIKELIYKTIQKLADEFKKNNNFVGIADLSTDYNH 226
Db      1495 IDDYILNKKAKINDCNVEKDEAHVKITKSLDKAIDDKIDLFKNTNDEFAIKKLINDDTK 1554
QY      227 NNLTKFLSGMWFEENLLKSTLSNLLDWKILARYVYKHFTTPMRKKTMTIOOQSGCPRHLDER 286
Db      1555 KDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISOHQCVKQCPENSGCPRHLDER 1612
QY      287 BECKCLLNYKQEGSKCVENPNPTCENNGGCDADAKTEEDSGSGKKITCOCTKPDSP 346
Db      1613 BECKCLLNYKQEGDKCVENPNPTCENNGGCDADATCTEEDSGSRKKITCECHKPDSP 1672
QY      347 LSMVIFCSSNFLGISFLLILMLILYSFI 375
Db      1673 LFDGIFCSSNFLGISFLLILMLILYSFI 1701

RESULT      9
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C:Accession: A26868
R:Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 199, 273-287, 1987
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum
A:Reference number: A26868; MUID:88011243
A:Accession: A26868
A:Molecule type: DNA
A:Residues: 1-1701 <VAV>
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match      54.8%; Score 1077; DB 2: Length 1701;
Best Local Similarity 55.5%; Pred. No. 3.8e-51;
Matches 217; Conservative 60; Mismatches 86; Indels 28; Gaps 5;

QY      9 NILSKIENEYEVLYLKLPLAGVYRSKKQLENNVMTFNVNKDLINSPFNKFNKLVLES 68
Db      1315 NILSGFENEVDVYLKPLAGVYRSKKQIEKNIITFNLNLDILNSLRKKRYFDLVLES 1374
QY      69 DLIPYKDLTSSNYVVKPYKFLNKEKDKLSSYNYIKDSIDTDINFANDVLGYKYLSE 128
Db      1375 DLMOFKHISNEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGISYIEKVL 1434
QY      129 KYKSDLSISIKYI-----NDKOGENEKYLFPFLNNIETLYKTVNDK 168
Db      1435 KYKDDLESIKKVIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETLYNNLVNK 1494
QY      169 IDLPIVHLEAKVLNVTYEK--SNVEIKELIYKTIQKLADEFKKNNNFVGIADLSTDYNH 226
Db      1495 IDDYILNKKAKINDCNVEKDEAHVKITKSLDKAIDDKIDLFKNTNDEFAIKKLINDDTK 1554
QY      227 NNLTKFLSGMW--FENLLKSTLSNLLDWKILARYVYKHFTTPMRKKTMTIOOQSGCPRHL 284
Db      1555 KDMLGKLLSTGLVQIFPN---TIISKIEGKF-QDMLNISOHQCVKQCPENSGCPRHL 1610
QY      285 ERECKCLLNYKQEGSKCVENPNPTCENNGGCDADAKCTEEDSGSGKKITCOCTKPDSP 344
Db      1611 ERECKCLLNYKQEGDKCVENPNPTCENNGGCDADATCTEEDSGSRKKITCECHKPDSP 1670

```


C:Genetics:
A:Note: MSV152

```
Query Match      8.4%; Score 165.5; DB 2; Length 1306;
Best Local Similarity 21.1%; Pred. No. 0.12;
Matches 72; Conservative 63; Mismatches 111; Indels 95; Gaps 13;

Qy 8 HNILSKTENEVEVLYLPLAGVYRSKQLENNVMTFNVNVDILNSPFNKRENFKNV-- 65
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 89 HNLQNIINNKTEI-----LPDELSFFNYNTANYLRFRMYDDNSSLRG 132

Qy 66 -----LESDLIPYKDLTSSNVVYKDPYKF---LNKEKRDKFLSSYNYIKDSIDTDINFA 116
   || || || || || || || || || || || || || || || || || || || ||
Db 133 GMLIDKTEDGFISYND-ESASYIVKKMENYISINEDKYDFYVYHAFI-----DYFLE 185

Qy 117 NDVLGYKYKILSEKYKSDL-----DSIKKYINDKOGENE----- 149
   || || || || || || || || || || || || || || || || || || || ||
Db 186 NDKLSYGDIDIDEKSKNDKMYNKDEYKKYLNEHTVSNDESKKIRKKIKYYLKFDTILHK 245

Qy 150 -----KYLPELNNIETLYKTVDNKIDLFVHLEAK-VLNYTYEKSNEVEIKELIYLKTIQ 202
   :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 246 DIISFEYIPLTKILDYIDKDTIDIKLIDLITITRAEPLN---ENNSTEIAYLFNKKQS 302

Qy 203 DKLADFRK-----NNN-----FVGTDLSTDYNNHNNLITKFLS--TGMVF 240
   :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:| :||:|
Db 303 LRLYENKKYNNNNNALSCKFVSNKLLDIOQRRIINIIMEINSDDNPTEILITNSKYQSII 362

Qy 241 ENLKSIL--SNLIDWKIARYVKHFTTPMRKKTKWIOQNSGC 279
   ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 363 NNINYAFITFTNYINNVLYVEFNLDSPNVKKFKIKFNCLC 403
```

Search completed: August 8, 2001, 12:30:58
Job time: 106 sec

THIS PAGE BLANK (USTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:58 ; Search time 34.41 Seconds
(without alignments)
224.393 Million cell updates/sec

Title: US-09-500-376-4

Perfect score: 1967

Sequence: 1 AVTPSVIHNLKSIENEYEV.....SNFLGISFLLILMLILYSFI 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCRUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	272	13.8	53	1	US-08-290-919-3
2	253	12.9	53	1	US-08-290-919-4
3	232	11.8	106	1	US-08-290-919-11
4	200	10.2	48	1	US-08-290-919-1
5	200	10.2	48	1	US-08-290-919-12
6	197	10.0	48	1	US-08-290-919-2
7	127	6.5	984	1	US-08-257-073-3
8	127	6.5	984	2	US-08-184-009-120
9	127	6.5	984	2	US-08-458-356-120
10	117.5	6.0	2391	2	US-08-446-855A-2
11	117.5	6.0	2391	4	US-09-150-741-2
12	117	5.9	3135	1	US-08-323-170B-2
13	114	5.8	1786	4	US-08-373-462-8
14	113	5.7	455	2	US-08-392-625-21
15	113	5.7	455	2	US-08-466-961A-21
16	113	5.7	455	2	US-08-645-193B-23
17	113	5.7	978	2	US-08-415-593-43
18	112.5	5.7	1147	3	US-08-470-260-5
19	112.5	5.7	1147	3	US-08-471-491-5
20	112.5	5.7	1147	4	US-08-466-662-5
21	112.5	5.7	2710	1	US-08-480-604A-6
22	112.5	5.7	2710	2	US-08-405-496A-6
23	112.5	5.7	3289	2	US-08-477-451-2
24	112	5.7	912	2	US-08-951-871-2
25	111.5	5.7	599	2	US-08-910-551B-2
26	109.5	5.6	1435	2	US-08-568-459A-4
27	109.5	5.6	1435	2	US-08-487-826B-4

28 108.5 5.5 720 2 US-08-840-236-1 Sequence 1, Appli
29 108.5 5.5 720 2 US-08-505-448A-1 Sequence 1, Appli
30 108.5 5.5 990 2 US-08-392-625-20 Sequence 20, Appli
31 108.5 5.5 990 2 US-08-466-961A-20 Sequence 20, Appli
32 107.5 5.5 990 2 US-08-645-193B-15 Sequence 15, Appli
33 107 5.4 1264 1 US-07-789-915A-6 Sequence 6, Appli
34 107 5.4 1264 1 US-08-005-002C-6 Sequence 6, Appli
35 107 5.4 1264 1 US-08-487-203A-6 Sequence 6, Appli
36 105.5 5.4 615 1 US-08-484-105-6 Sequence 6, Appli
37 105.5 5.4 615 1 US-08-484-106-6 Sequence 6, Appli
38 105 5.3 934 3 US-08-929-329-6 Sequence 6, Appli
39 105 5.3 1579 3 US-08-755-587-184 Sequence 184, App
40 103.5 5.3 730 4 US-09-398-865A-2 Sequence 2, Appli
41 103 5.2 1248 2 US-09-080-897-2 Sequence 2, Appli
42 103 5.2 1248 4 US-09-323-735-2 Sequence 2, Appli
43 103 5.2 1255 2 US-09-080-897-4 Sequence 4, Appli
44 103 5.2 1255 3 US-08-899-595-1 Sequence 1, Appli
45 103 5.2 1255 4 US-09-323-735-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-290-919-3
; Sequence 3, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714827 CUSH
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site

```

; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X = M and N, or N"
US-08-290-919-3

```

Query Match	13.8%	Score 272;	DB 1;	Length 53;
Best Local Similarity	92.3%	Pred. No. 2.9e-15;		
Matches 48;	Conservative	1;	Mismatches 3;	Indels 0

2008 Local Community Survey, 2008
 Matches 48; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy	306	PNPTC	NENNG	GDAD	AKCTE	DGSGN	GKKIT	CQCTK	PDSP	PLSM	VIFC	SSN	357
Db	2	PNPTC <th>NENNG</th> <th>GDAD</th> <th>AKCTE</th> <th>DGSGN</th> <th>GKKIT</th> <th>CCTCK</th> <th>PDSP</th> <th>PLFD</th> <th>GIFC</th> <th>SSN</th> <th>53</th>	NENNG	GDAD	AKCTE	DGSGN	GKKIT	CCTCK	PDSP	PLFD	GIFC	SSN	53

RESULT 2

US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918

[illegible]

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
;

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 4:

```

;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 53 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; FEATURE:

```

```

; NAME/KEY: Modified-site
; LOCATION: 1

```

```

; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X = M and N, or N"

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Query Match	12.9%;	Score 253;	DB 1;	Length 53;
Best Local Similarity	86.5%;	Pred. No. 9.3e-14;		

		Matches	45;	Conservative	2;	Mismatches	5;	Indels	0;	Gaps	0;
Qy	306	PNPTCNENGGCDADAKCTEEDSGSGKKITCQCTKPDSPLSMVFCS	SN	357							
Db	2	PNPTCNENGGCDADAKCTEEDSGSGKKITCCTKPDSPYPLDGFCS	SN	53							

RESULT 3

US-08-230-919-11
; Sequence 11, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER; ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.

COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9203821.5
 ; FILING DATE: 22-FEB-1992

, FILING DATE. 22 FEB 1992
 ;
 ; PRIOR APPLICATION DATA:
 : APPLICATION NUMBER. PCT/CP92/00367

; APPLICATION NUMBER: PCI/GB93/00367
 ; FILING DATE: 22-FEB-1993
 ; ATTORNEY / AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16 773

REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/

TELECOMMUNICATION INFORMATION
TELEPHONE: (202) 861-3000

TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH

```

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 106 amino acids
; TYPE:  amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
US-08-290-919-11

Query Match	11.8%
Best Local Similarity	48.0%

Matches 47; Conservative 11;

Qy 261 KHETPMRKKTMIIQQNSGCFRHLDP

Db 12 KHVCVDTRD--IPKNAGCFRDDNC

QY 320 DAKTEEDSGSNGKKITCQCTKPD

Db 69 TASCQNAESTENSKKIICTCKEPTN

E
 H
 C
 D
 C

RESULT 4
US-08-290-919-1


```
; Sequence 1, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/NJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X = M and N, or N"
US-08-290-919-1

Query Match 10.2%; Score 200; DB 1; Length 48;
Best Local Similarity 87.5%; Pred. No. 1.4e-09;
Matches 35; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 270 KTMIQNSGCFRHLDERECKLLNYKQEGSKVCVENPNPT 309
Db 9 KKQCPNSGCFRHLDERECKLLNYKQEGDKVCVENPNPT 48

RESULT 5
US-08-290-919-12
; Sequence 12, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/NJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-290-919-12

Query Match 10.2%; Score 200; DB 1; Length 48;
Best Local Similarity 87.5%; Pred. No. 1.4e-09;
Matches 35; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 270 KTMIQNSGCFRHLDERECKLLNYKQEGSKVCVENPNPT 309
Db 9 KKQCPNSGCFRHLDERECKLLNYKQEGDKVCVENPNPT 48

RESULT 6
US-08-290-919-2
; Sequence 2, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELETYPE: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= X
; OTHER INFORMATION: /note= "X = M and N, or N"
US-08-290-919-2

Query Match 10.08; Score 197; DB 1; Length 48;
Best Local Similarity 85.08; Pred. No. 2.4e-09;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 270 KTMQNGSGCFRHLDERECKLLNKOEGSKVCVENPNPT 309
DB 9 KQCPENSGCFRHLDERECKLLNKOEGSKVCVENPNPT 48

RESULT 7
US-08-257-073-3
; Sequence 3, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
```

```
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELETYPE: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-257-073-3

Query Match 6.5%; Score 127; DB 1; Length 984;
Best Local Similarity 20.4%; Pred. No. 0.051;
Matches 94; Conservative 64; Mismatches 161; Indels 142; Gaps 21;

QY 4 PSVIHNLISKIENEYEVLYKPLA-----GVYRSLKKOLENN---VMTFNVNVKDIILNSP 55
DB 236 PRNLQNICETGCKFKLVYIKENTLIHKVYGETKDTTENNKKVDVRYKLYLNEKE---TP 292
QY 56 FN-----KRENFKNVLES-----DLIPVK-DLTSSNVVVKDPYKF-----LNKEK 94
DB 293 FTSILIHAYKEHNGTNIIESKNYALGSDIPEKCDTLASNCFLSGNFNIEKCFQCALIVEK 352
QY 95 RDKFLSSYNIKDSIDTDFANDVLGYKILSEKYSKDLDSIKKYINDKQGENEYKLPF 154
DB 353 ENK-----NDVC--YKVLSEDIYSKFEKAEDEDDDDDDYTEYKL 391
QY 155 LNNIET---LYKT--VNDKIDLFVIH-----LEAKVLNITYEKSNEVEIKELIYLKTIOD 203
DB 392 TESIDNILVKMKTNNENKSELIKLEEVDDSLKLEMLNYCSSLKLDVDTTGTLDNYGMGN 451
QY 204 KLADFKKNNFVGIADLSTDYNNHNNLTKFLSTGMVFENL-----LKSILSNLLDWKLA 257
DB 452 EMDIF---NNLKRLLIYHSEENINTLANKFENAAVCLKNVDDWIVNKRGLVLPENLYDLE 508
QY 258 RYVKHF-----TPMRKKTMTIQNS-----GC 279
DB 509 YFNEHLYNDKNSPEDKNGKGVVHVDTTLEKEDTLSYDSNMFNFCNKEYCNRLKDNNC 568
QY 280 FRHLDERECKC-----LLNYKQEGSKVCVENPNPT-----C--NENNGCDADAK 322
DB 569 ISNLQVEDQNCNCDTSWIFASKYHLETRCMKGYEPTKISALYVANCYKGEHKRCDGSGS 628
QY 323 CTE-----EDSG-----SNGKKITCQCTKPDSPSLM 349
DB 629 PMEFLQIIEDYGFPAESNYPYNYVKVGEQCPKVEDHMMNL 669

RESULT 8
US-08-184-009-120
; Sequence 120, Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
```

CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/184,009
FILING DATE: 19-JAN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-0712
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-184-009-120

Query Match 6.5%; Score 127; DB 2; Length 984;
Best Local Similarity 20.4%; Pred. No. 0.051;
Matches 94; Conservative 64; Mismatches 161; Indels 142; Gaps 21;

Qy 4 PSVTHNLSKIENEYEVLYLKPLA-----GYVRSLLKQLENN---VMTFNVNVDILNSP 55
Db 236 PRNLQNICETGKNEFLVYIKENTLIWKVYGETKTOTENNKVDVRKYLINEKE---TP 292
Qy 56 FN-----KRENPKNVLES-----DLIPYK-DLTSNYSVVKDPYK-----LNKEK 94
Db 293 FTSILIHAYKEHNGTNLIESKNYALGSDIPEKCDTLASNCFLSGNFNIEKCFQCALLIVEK 352
Qy 95 RDKFLLSYNYIKDSIDTDINFANDVLGYKILSEKYSKLDLSIKKYINDKOGENEKYLPPF 154
Db 353 ENK-----NDVC--YKYLSEDIIVSKFEIKAETDEDDDDYTEYKL 391
Qy 155 LNNTET-----LYKT--VNDKIDLFVH-----LEAKVLNITYERSNVEIKEIYLKTIQD 203
Db 392 TESIDNVLVKMFTNENNDKSELIKLEEVDDSLKLELMNYCSLLKDVDTTCTLDNYGMGN 451
Qy 204 KLADFKNNNPFVGIADLSTDYNNHNLTKFLSTGCVFENL-----LKSILSNLLDWKLA 257
Db 452 EMDIF---NNLKRLLIYHSEENINTLNKFRNAACVCLKNVDVIVNKRGLVLPDLNVDLE 508
Qy 258 RYVKHF-----TPMRKKTMTIQONS-----GC 279
Db 509 YFNEHLYNDKNSPEDKDKNGKGVVHVDTTLEKEDTLSDYNSDNMFCNKCYNRLKDNNC 568
Qy 280 FRHLDERECKC-----LLNYKQEGSKCVENPNPT-----C--NENNGCCDADAK 322
Db 569 ISNLOVEDOGNCDSWIFASKYHLETRICMKGYEPTKISALYVANCYKGEHKDCDEGSS 628
Qy 323 CTE-----EDSG-----SNGKKITCQCTKPDSPLSM 349
Db 629 PMEFLOIYEDYGLFAESNYPYNTVYKVGEOCPKVEDHWMNL 669

RESULT 9

US-08-458-356-120
; Sequence 120, Application US/08458356
; Patent No. 5942235

GENERAL INFORMATION:
APPLICANT: Paolletti, Enzo
APPLICANT: Tartaglia, James
APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,356
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/184,009
FILING DATE: 19-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 120:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-458-356-120

Query Match 6.5%; Score 127; DB 2; Length 984;
Best Local Similarity 20.4%; Pred. No. 0.051;
Matches 94; Conservative 64; Mismatches 161; Indels 142; Gaps 21;

Qy 4 PSVTHNLSKIENEYEVLYLKPLA-----GYVRSLLKQLENN---VMTFNVNVDILNSP 55
Db 236 PRNLQNICETGKNEFLVYIKENTLIWKVYGETKTOTENNKVDVRKYLINEKE---TP 292
Qy 56 FN-----KRENPKNVLES-----DLIPYK-DLTSNYSVVKDPYK-----LNKEK 94
Db 293 FTSILIHAYKEHNGTNLIESKNYALGSDIPEKCDTLASNCFLSGNFNIEKCFQCALLIVEK 352
Qy 95 RDKFLLSYNYIKDSIDTDINFANDVLGYKILSEKYSKLDLSIKKYINDKOGENEKYLPPF 154
Db 353 ENK-----NDVC--YKYLSEDIIVSKFEIKAETDEDDDDYTEYKL 391
Qy 155 LNNTET-----LYKT--VNDKIDLFVH-----LEAKVLNITYERSNVEIKEIYLKTIQD 203
Db 392 TESIDNVLVKMFTNENNDKSELIKLEEVDDSLKLELMNYCSLLKDVDTTCTLDNYGMGN 451
Qy 204 KLADFKNNNPFVGIADLSTDYNNHNLTKFLSTGCVFENL-----LKSILSNLLDWKLA 257
Db 452 EMDIF---NNLKRLLIYHSEENINTLNKFRNAACVCLKNVDVIVNKRGLVLPDLNVDLE 508
Qy 258 RYVKHF-----TPMRKKTMTIQONS-----GC 279
Db 509 YFNEHLYNDKNSPEDKDKNGKGVVHVDTTLEKEDTLSDYNSDNMFCNKCYNRLKDNNC 568
Qy 280 FRHLDERECKC-----LLNYKQEGSKCVENPNPT-----C--NENNGCCDADAK 322

Db 569 ISNLOVEDOQCDTSWIFASKYHLETRCMKGYEPTKRISALYVANCYKGBKDRCDGSS 628
QY 323 CTE-----EBSG-----SNGKKTKTCQCTRPDSPLSM 349
Db 629 PMEFLOIIEDYGLPAESNYPYVVKYGEQCPKVEDHWMNL 669
RESULT 10
US-08-446-855A-2
; Sequence 2, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 NO. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446, 855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29,009
; REFERENCE/DOCKET NUMBER: 47-80
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-446-855A-2

Query Match 6.0%; Score 117.5; DB 2; Length 2391;
Best Local Similarity 20.1%; Pred. No. 0.95;
Matches 69; Conservative 60; Mismatches 111; Indels 103; Gaps 16;
QY 37 LENNV---MTFNVNVKDLSPNFKNFVLESOLIPYKDLTSSNVV-----KD 85
Db 1333 IENNICHDISINKIKYTIINNSNISNNENV-ETNLCVSRAGSHHIYKKEKSGSD 1391
QY 86 PYKFLNKKRDKFLSSVNYIKDSITDIDIN-FANDVLGYKILSEKYSKDLSDSIKKY---- 140
Db 1392 DTNLSAQNSNNFSCNENNNKANKVDVNVLENDT-----KKREDINTTTFVMEGQ 1442
QY 141 ---INDKQGENEKYLPFLNNIETLYKTVDNKIDLFVHLEAKVLYNTEKSNVEI-KELI 196
Db 1443 NSVINNNKENSILK-----GDEEDIVMVNLIK-KENNYNSVINNVDCRKKDM 1489
QY 197 YLKTIOQKLADEFK-----NNNFVGIADLSTDYHNHNLTKFLSTGMVFENLLKSIL 248
Db 1490 DGKNIINDECKTYKKNYKDKMGLNNIVDELNSGTSHSTNDHL----- 1531
QY 249 SNLLDWKLARYVKHFTTP-----MRKKTMIQOQSG-CFRHLDERECKCLL 293
Db 1532 ---YLDNFNTSDEEIGNKKNMDMYLSKEKSIKKNFGNSYVYVDSVYNNEYKI 1581
QY 294 NYKQBSKCVENPNPTCENNNGGCDADAKCTEEDSGS---NGK 333
Db 1582 NKMK-----LIDNENLNDEYNN---NVNMNCSNYNNASAFVNGK 1618

Db 1532 -----YLDNFNTSDEEIGNKKNMDMYLSKEKSIKKNFGNSYVYVDSVYNNEYKI 1581
QY 294 NYKQBSKCVENPNPTCENNNGGCDADAKCTEEDSGS---NGK 333
Db 1582 NKMK-----LIDNENLNDEYNN---NVNMNCSNYNNASAFVNGK 1618
RESULT 11
US-09-150-741-2
; Sequence 2, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2391
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: protein
; US-09-150-741-2

Query Match 6.0%; Score 117.5; DB 4; Length 2391;
Best Local Similarity 20.1%; Pred. No. 0.95;
Matches 69; Conservative 60; Mismatches 111; Indels 103; Gaps 16;
QY 37 LENNV---MTFNVNVKDLSPNFKNFVLESOLIPYKDLTSSNVV-----KD 85
Db 1333 IENNICHDISINKIKYTIINNSNISNNENV-ETNLCVSRAGSHHIYKKEKSGSD 1391
QY 86 PYKFLNKKRDKFLSSVNYIKDSITDIDIN-FANDVLGYKILSEKYSKDLSDSIKKY---- 140
Db 1392 DTNLSAQNSNNFSCNENNNKANKVDVNVLENDT-----KKREDINTTTFVMEGQ 1442
QY 141 ---INDKQGENEKYLPFLNNIETLYKTVDNKIDLFVHLEAKVLYNTEKSNVEI-KELI 196
Db 1443 NSVINNNKENSILK-----GDEEDIVMVNLIK-KENNYNSVINNVDCRKKDM 1489
QY 197 YLKTIOQKLADEFK-----NNNFVGIADLSTDYHNHNLTKFLSTGMVFENLLKSIL 248
Db 1490 DGKNIINDECKTYKKNYKDKMGLNNIVDELNSGTSHSTNDHL----- 1531
QY 249 SNLLDWKLARYVKHFTTP-----MRKKTMIQOQSG-CFRHLDERECKCLL 293
Db 1532 ---YLDNFNTSDEEIGNKKNMDMYLSKEKSIKKNFGNSYVYVDSVYNNEYKI 1581
QY 294 NYKQBSKCVENPNPTCENNNGGCDADAKCTEEDSGS---NGK 333
Db 1582 NKMK-----LIDNENLNDEYNN---NVNMNCSNYNNASAFVNGK 1618

RESULT 12
US-08-323-170B-2
; Sequence 2, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.

```

RESULT 14
US-08-392-625-21
; Sequence 21, Application US/08392625
; Patent No. 5837485
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
; APPLICANT: G tz, Friedrich
; APPLICANT: Schnell, No. 5837485bert
; APPLICANT: Augustin, Johannes
; APPLICANT: Engelke, Gernar
; APPLICANT: Rosenstein, Ralf
; APPLICANT: Kaletta, Cortina
; APPLICANT: Klein, Cora
; APPLICANT: Wieland, Bernd
; APPLICANT: Kupke, Thomas
; APPLICANT: Jung, G nther
; APPLICANT: Kellner, Roland
; TITLE OF INVENTION: Biosynthetic Process For The Preparation
; OF Chemical Compounds
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA: 0652.09800002
 APPLICATION NUMBER: US/08/392,625
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/876,791
 FILING DATE: 30-APR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Esmond, Robert W.
 REGISTRATION NUMBER: 32,893
 REFERENCE/DOCKET NUMBER: 0652.09800002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 371-2600
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 455 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-392-625-21

[illegible]

RESULT 15
US-08-466-961A-21
; Sequence 21, Application US/08466961A
; Patent No. 5843709
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
; APPLICANT: G tz, Friedrich
; APPLICANT: Schnell, No. 5843709bert
; APPLICANT: Augustin, Johannes
; APPLICANT: Engelke, Germar
; APPLICANT: Rosenstein, Ralf
; APPLICANT: Kaletta, Cortina
; APPLICANT: Klein, Cora
; APPLICANT: Wieland, Bernd
; APPLICANT: Kupke, Thomas
; APPLICANT: Jung, G nther
; APPLICANT: Kellner, Roland
; TITLE OF INVENTION: Biosynthetic Process for the Preparation of
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

```

: STREET: 1100 New York Avenue, NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,961A
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/392,625
: FILING DATE: 22-FEB-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/876,791
: FILING DATE: 30-APR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/784,234
: FILING DATE: 31-OCT-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: REFERENCE/DOCKET NUMBER: 0652.0980004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 455 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-08-466-961A-21

Query Match 5.7%; Score 113; DB 2; Length 455;
Best Local Similarity 22.2%; Pred. No. 0.24;
Matches 80; Conservative 51; Mismatches 119; Indels 110; Gaps

Qy 17 EYEVLYLKPLAGVYRSKK--QLEN-NVMTFNNVKNVDILNSPFKNRENFKNVLESLLPY 73
Db 11 EYVLLKKNLFLFLMKLQKLNIGVYININ- - - - - NIKILLENKITFL 57
Qy 74 KDLTSSNYVK-----DPYK-----FLNKEKR-----DKFLSSY-----NYIK 106
Db 58 SDIEKATVIENSEQYWDPTLSHGYPGILFLSASEKVFHKDLEKVHQYIRKLGPYLE 117
Qy 107 DSID-----TDINFANDVLGYKKILSEK---YKSDILDSIK----KYINDKQGENEY 151
Db 118 SGIDGFSFLSGLSIGFALD-----IASDKQVSYSQILEQIDNLLVQYVFD----- 163
Qy 152 LPFLNN-----IEFLYKTVNDKIDFLVHLEAKVLNVTYEKSNVEIKELIYKTIQDKLA 206
Db 164 --FLNNDALEWTPPNYDIIQGFSGVGYLYLNRISYNNAKKALKHI--LNYFKTIH----- 215
Qy 207 DFKKNNFVGIADLSTDYNNHNNLTKFLSTGM---VFENLLKSLNSLLDWKL---ARYV 260
Db 216 -YSKDNVLVSNHEHQFLDIDKQFPFSGNINLGLAHGILGSLTALSCKMNGIEIEGHEEFL 274
Qy 261 KHFTPTMRKKTMIQONGGCFPHLDERECKLLNFKQEGSKCVENPNPTCENNNGGCDAD 320
Db 275 QDFTSFLLKPB-FKNNNEWDFRD-----ILENYIPNYSVRNGWCYGD 316

```

Search completed: August 8, 2001, 12:30:01
Job time: 49 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:33:42 ; Search time 70.07 Seconds
(without alignments)
743.944 Million cell updates/sec

Title: US-09-500-376-3

Perfect score: 2052

Sequence: 1 AISVTMDNLGFEENEYDVI.....SNFLGTSFLILMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_unclassified.*
- 13: sp_vertebrate.*
- 14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2033	99.1	570	5	Q25968
2	2033	99.1	570	5	Q9TYG2
3	2017.5	98.3	569	5	Q25970
4	2017.5	98.3	569	5	Q25980
5	2017.5	98.3	569	5	Q25982
6	2015	98.2	1720	5	Q25922
7	2013.5	98.1	569	5	Q25978
8	2011	98.0	652	5	Q25923
9	2001.5	97.5	569	5	Q25969
10	2001.5	97.5	569	5	Q25974
11	2001.5	97.5	569	5	Q25975
12	2001.5	97.5	569	5	Q25977
13	2001.5	97.5	569	5	Q25979
14	2001.5	97.5	1694	5	Q9TZT5
15	2001.5	97.5	1694	5	Q9NHX1
16	2001.5	97.5	1704	5	Q9TZT4
17	1997.5	97.3	569	5	Q25983
18	1987.5	96.9	651	5	Q25924
19	1982.5	96.6	569	5	Q25967

20	1930	94.1	373	5	Q25727	Q25727 plasmidium
21	1925	93.8	373	5	Q25728	Q25728 plasmidium
22	1919.5	93.5	372	5	Q25717	Q25717 plasmidium
23	1914.5	93.3	372	5	Q25718	Q25718 plasmidium
24	1914.5	93.3	372	5	Q25719	Q25719 plasmidium
25	1914.5	93.3	372	5	Q25720	Q25720 plasmidium
26	1910	93.1	373	5	Q43996	Q43996 plasmidium
27	1905	92.8	373	5	Q25724	Q25724 plasmidium
28	1900	92.6	373	5	Q43995	Q43995 plasmidium
29	1900	92.6	373	5	Q25721	Q25721 plasmidium
30	1898	92.5	373	5	Q25723	Q25723 plasmidium
31	1891	92.2	373	5	Q25722	Q25722 plasmidium
32	1887.5	92.0	372	5	Q25725	Q25725 plasmidium
33	1882.5	91.7	372	5	Q25726	Q25726 plasmidium
34	1881.5	91.7	372	5	Q43997	Q43997 plasmidium
35	1257.5	61.3	539	5	Q25972	Q25972 plasmidium
36	1257.5	61.3	539	5	Q25981	Q25981 plasmidium
37	1254.5	61.1	400	5	Q03999	Q03999 plasmidium
38	1254.5	61.1	539	5	Q25966	Q25966 plasmidium
39	1254.5	61.1	539	5	Q25976	Q25976 plasmidium
40	1254.5	61.1	539	5	Q25984	Q25984 plasmidium
41	1254.5	61.1	539	5	Q9TYG1	Q9TYG1 plasmidium
42	1248.5	60.8	539	5	Q25971	Q25971 plasmidium
43	1248.5	60.8	539	5	Q25973	Q25973 plasmidium
44	1142	55.7	218	5	Q9TVG8	Q9TVG8 plasmidium
45	828	40.4	1726	5	Q02569	Q02569 plasmidium

ALIGNMENTS

RESULT	1					
Q25968						
ID	Q25968	PRELIMINARY:	PRT:	570 AA.		
AC	Q25968;					
DT	01-NOV-1996 (TReMBLrel. 01, Created)					
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)					
DT	01-MAY-2000 (TReMBLrel. 13, Last annotation update)					
DE	MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).					
GN	MSPL.					
OS	Plasmodium falciparum.					
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.					
OX	NCBI_TaxID=5833;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=93295445; PubMed=8515786;					
RA	Jongwutives S., Tanabe K., Kanbara H.;					
RT	"Sequence conservation in the C-terminal part of the precursor to the					
RT	major merozoite surface proteins (MSPL) of Plasmodium falciparum from					
RT	field isolates";					
RL	Mol. Biochem. Parasitol. 59:95-100(1993).					
DR	EMBL; D13345; BAA02606.1; -					
DR	InterPro; IPR000561; -					
DR	Pfam; PF00008; EGF; 1.					
KW	Merozoite; EGF-like domain.					
FT	NON_TER					
SQ	SEQUENCE 570 AA; 64632 MW; 424BF553CCC2F2BE CRC64;					
Query Match 99.1%; Score 2033; DB 5; Length 570;						
Best Local Similarity 99.2%; Pred. No. 2.3e-101;						
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;						
Qy	1	AISVTMDNLGFEENEYDVIYKPLAGVYRSLLKQIEKNIITFNLNLDILNSLRKKRY	60			
Db	177	AISVTMDNLGFEENEYDVIYKPLAGVYRSLLKQIEKNIITFNLNLDILNSLRKKRY	236			
Qy	61	FLDVLSDLMQFKHISSEYIIIESFKLLNSEQNKLLSKYIKESVENDIKFAQEGIS	120			
Db	237	FLDVLSDLMQFKHISSEYIIIESFKLLNSEQNKLLSKYIKESVENDIKFAQEGIS	296			
Qy	121	YIEKVLAKYKDDLESIKVKIIEKEKEKPPSPPTTTPPSAKTDEQKESKFLPFTLTNETL	180			

```
Db 297 YEKVLAKYKDDLESIKVKIEKEKFPSPPTTPPAKTDEQKESKFLPFLTNIETL 356
QY 181 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 357 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 416
QY 241 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEGFQFMDLNIHQCVKQKCPENSGCFR 300
Db 417 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLEGFQFMDLNIHQCVKQKCPENSGCFR 476
QY 301 HLDRECKCLLNYKQEGDKCEENPNPTCNENNGCGDADATCTEEDSGSSRKKITCETCK 360
Db 477 HLDRECKCLLNYKQEGDKCEENPNPTCNENNGCGDADATCTEEDSGSSRKKITCETCK 536
QY 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db 537 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 570
RESULT 2
ID Q9TYG2 PRELIMINARY; PRT; 570 AA.
AC Q9TYG2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum."
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13343; BAA02604.1; -
DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 570 AA; 64630 MW; 8674DEC89B2D662A CRC64;
```

```
Query Match 99.18; Score 2033; DB 5; Length 570;
Best Local Similarity 99.28; Pred. No. 2.3e-101;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AISVTMDNILSGFENEYDVLYLPLAGVYSLKKQIEKNITFNLDNLINSLRKKRY 60
Db 177 AISVTMDNILSGFENEYDVLYLPLAGVYSLKKQIEKNITFNLDNLINSLRKKRY 236
QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 296
QY 121 YEKVLAKYKDDLESIKVKIEKEKFPSPPTTPPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKVKIEKEKFPSPPTTPPAKTDEQKESKFLPFLTNIETL 356
QY 181 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 357 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 416
```

```
QY 241 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEGFQFMDLNIHQCVKQKCPENSGCFR 300
Db 417 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLEGFQFMDLNIHQCVKQKCPENSGCFR 476
QY 301 HLDRECKCLLNYKQEGDKCEENPNPTCNENNGCGDADATCTEEDSGSSRKKITCETCK 360
Db 477 HLDRECKCLLNYKQEGDKCEENPNPTCNENNGCGDADATCTEEDSGSSRKKITCETCK 536
QY 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db 537 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 570
RESULT 3
ID Q25970 PRELIMINARY; PRT; 569 AA.
AC Q25970;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13347; BAA02608.1; -
DR InterPro: IPR000561; -
DR Pfam: PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64502 MW; 4D761FF472753142 CRC64;
Query Match 98.38; Score 2017.5; DB 5; Length 569;
Best Local Similarity 99.08; Pred. No. 1.5e-100;
Matches 390; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 1 AISVTMDNILSGFENEYDVLYLPLAGVYSLKKQIEKNITFNLDNLINSLRKKRY 60
Db 177 AISVTMDNILSGFENEYDVLYLPLAGVYSLKKQIEKNITFNLDNLINSLRKKRY 236
QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 296
QY 121 YEKVLAKYKDDLESIKVKIEKEKFPSPPTTPPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKVKIEKEKFPSPPTTPPAKTDEQKESKFLPFLTNIETL 355
QY 181 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 356 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
QY 241 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEGFQFMDLNIHQCVKQKCPENSGCFR 300
Db 416 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLEGFQFMDLNIHQCVKQKCPENSGCFR 475
QY 301 HLDRECKCLLNYKQEGDKCEENPNPTCNENNGCGDADATCTEEDSGSSRKKITCETCK 360
Db 476 HLDRECKCLLNYKQEGDKCEENPNPTCNENNGCGDADATCTEEDSGSSRKKITCETCK 535
QY 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569
```

```
RESULT 4
Q25980 ID Q25980 PRELIMINARY; PRT; 569 AA.
AC Q25980;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP MEDLINE-93295445; PubMed-8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13353; BAA02614.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
FT SEQUENCE 569 AA; 64536 MW; 8008861DECECD8DC CRC64;

Query Match 98.3%; Score 2017.5; DB 5; Length 569;
Best Local Similarity 99.0%; Pred. No. 1.5e-100;
Matches 390; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1 AISTVMDNLSGFENEYDVIYKPLAGYRSKQIEKNIITFNLDILNSRLKRRKY 60
Db 177 AISTVMDNLSGFENEYDVIYKPLAGYRSKQIEKNIITFNLDILNSRLKRRKY 236
Qy 61 FLDVLESQPMQFKHSSNEYIIEDSFKLNSQKNIILSKYKESVENDIKFAQEGIS 120
Db 237 FLDVLESQPMQFKHSSNEYIIEDSFKLNSQKNIILSKYKESVENDIKFAQEGIS 296
Qy 121 YEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 355
Qy 181 YNNLVNKIDDDYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 356 YNNLVNKIDDDYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
Qy 241 LINDTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCYKQKCPENSGCGR 300
Db 416 LINDTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCYKQKCPENSGCGR 475
Qy 301 HLDERECKLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
Db 476 HLDERECKLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535
Qy 361 PDSYPLFGICFSSSNFLGISFLILMLILYSFI 394
Db 536 PDSYPLFGICFSSSNFLGISFLILMLILYSFI 569

RESULT 5
Q25982 ID Q25982 PRELIMINARY; PRT; 569 AA.
AC Q25982;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP MEDLINE-93295445; PubMed-8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13353; BAA02614.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
FT SEQUENCE 569 AA; 64536 MW; 8008861DECECD8DC CRC64;

Query Match 98.3%; Score 2017.5; DB 5; Length 569;
Best Local Similarity 99.0%; Pred. No. 1.5e-100;
Matches 390; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1 AISTVMDNLSGFENEYDVIYKPLAGYRSKQIEKNIITFNLDILNSRLKRRKY 60
Db 177 AISTVMDNLSGFENEYDVIYKPLAGYRSKQIEKNIITFNLDILNSRLKRRKY 236
Qy 61 FLDVLESQPMQFKHSSNEYIIEDSFKLNSQKNIILSKYKESVENDIKFAQEGIS 120
Db 237 FLDVLESQPMQFKHSSNEYIIEDSFKLNSQKNIILSKYKESVENDIKFAQEGIS 296
Qy 121 YEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 355
Qy 181 YNNLVNKIDDDYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 356 YNNLVNKIDDDYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
Qy 241 LINDTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCYKQKCPENSGCGR 300
Db 416 LINDTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCYKQKCPENSGCGR 475
Qy 301 HLDERECKLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
Db 476 HLDERECKLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535
Qy 361 PDSYPLFGICFSSSNFLGISFLILMLILYSFI 394
Db 536 PDSYPLFGICFSSSNFLGISFLILMLILYSFI 569

RESULT 6
Q25922 ID Q25922 PRELIMINARY; PRT; 1720 AA.
AC Q25922;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE OF 1069-1720 FROM N.A.
RC STRAIN=NF54;
RA Tolle R.; Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
```

```
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13354; BAA02615.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
FT SEQUENCE 569 AA; 64530 MW; 4D734580516D2142 CRC64;

Query Match 98.3%; Score 2017.5; DB 5; Length 569;
Best Local Similarity 99.0%; Pred. No. 1.5e-100;
Matches 390; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1 AISTVMDNLSGFENEYDVIYKPLAGYRSKQIEKNIITFNLDILNSRLKRRKY 60
Db 177 AISTVMDNLSGFENEYDVIYKPLAGYRSKQIEKNIITFNLDILNSRLKRRKY 236
Qy 61 FLDVLESQPMQFKHSSNEYIIEDSFKLNSQKNIILSKYKESVENDIKFAQEGIS 120
Db 237 FLDVLESQPMQFKHSSNEYIIEDSFKLNSQKNIILSKYKESVENDIKFAQEGIS 296
Qy 121 YEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKVKIKEEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 355
Qy 181 YNNLVNKIDDDYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 356 YNNLVNKIDDDYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
Qy 241 LINDTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCYKQKCPENSGCGR 300
Db 416 LINDTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCYKQKCPENSGCGR 475
Qy 301 HLDERECKLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 360
Db 476 HLDERECKLLNYKQEGDKCEENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK 535
Qy 361 PDSYPLFGICFSSSNFLGISFLILMLILYSFI 394
Db 536 PDSYPLFGICFSSSNFLGISFLILMLILYSFI 569

RESULT 6
Q25922 ID Q25922 PRELIMINARY; PRT; 1720 AA.
AC Q25922;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5843;
RN [1]
RP SEQUENCE OF 1069-1720 FROM N.A.
RC STRAIN=NF54;
RA Tolle R.; Bujard H., Cooper J.A.;
RL Exp. Parasitol. 0:0-0(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
```

SEQUENCE FROM N.A.
STRAIN=NF54;
MEDLINE=96123395; PubMed=8577332;
Pan W., Tolle R., Bujard H.;
"A direct and rapid sequencing strategy for the Plasmodium falciparum
antigen gene gp190/MSA1";
Mol. Biochem. Parasitol. 73:241-244(1995).
EMBL: Z35327; CAA84556.1; -
InterPro: IPR000561; -
Pfam: PF00008; EGF; 1.
Signal; Merozoite.
FT SIGNAL 1 19 POTENTIAL.
SQ SEQUENCE 1720 AA; 195725 MW; 71782FD1E637A8A3 CRC64;

Query Match 98.28; Score 2015; DB 5; Length 1720;
Best Local Similarity 98.58; Pred. No. 6.6e-100;
Matches 388; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIITFNLDILNSRLKRRKY 60
Db 1327 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIITFNLDILNSRLKRRKY 1386

Qy 61 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNKLLSKYIKESVENDIKFAQEGIS 120
Db 1387 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNKLLSKYIKESVENDIKFAQEGIS 1446

Qy 121 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 1447 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 1506

Qy 181 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLKFNNDPEAIKK 240
Db 1507 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLKFNNDPEAIKK 1566

Qy 241 LINDTKMDLGLKLLSTGLVQIFNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 300
Db 1567 LINDTKMDLGLKLLSTGLVQIFNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 1626

Qy 301 HLDRECKCLLNKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSSRKKITCECTK 360
Db 1627 HLDRECKCLLNKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSSRKKITCECTK 1686

Qy 361 PDSYPLFDGIFCSCSNFLGISFLILMLILYSFI 394
Db 1687 PDSYPLFDGIFCSCSNFLGISFLILMLILYSFI 1720

RESULT 7
Q25978 ID Q25978 PRELIMINARY; PRT; 569 AA.
AC Q25978
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RA Tolle R.;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
major merozoite surface proteins (MSP) of Plasmodium falciparum from
field isolates";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
EMBL: D13351; BAA02612.1; -
InterPro: IPR000561; -
Pfam: PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1 1

SQ SEQUENCE 569 AA; 64536 MW; B8B9B63EECBIDA51 CRC64;

Query Match 98.18; Score 2013.5; DB 5; Length 569;
Best Local Similarity 98.78; Pred. No. 2.5e-100;
Matches 389; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIITFNLDILNSRLKRRKY 60
Db 177 AISVTMDNLISGFENEYDVYILKPLAGVYRSLLKQIEKNIITFNLDILNSRLKRRKY 236

Qy 61 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNKLLSKYIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNKLLSKYIKESVENDIKFAQEGIS 296

Qy 121 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 355

Qy 181 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLKFNNDPEAIKK 240
Db 356 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLKFNNDPEAIKK 415

Qy 241 LINDTKMDLGLKLLSTGLVQIFNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 300
Db 416 LINDTKMDLGLKLLSTGLVQIFNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 475

Qy 301 HLDRECKCLLNKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSSRKKITCECTK 360
Db 476 HLDRECKCLLNKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSSRKKITCECTK 535

Qy 361 PDSYPLFDGIFCSCSNFLGISFLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSCSNFLGISFLILMLILYSFI 569

RESULT 8
Q25923 ID Q25923 PRELIMINARY; PRT; 652 AA.
AC Q25923
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RA Tolle R.;
RN [1]
RP SEQUENCE OF 1-298 FROM N.A.
RX STRAIN=FCHE5/NF7;
RA Peterson G.M., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
Brown G.V., Anders R.P., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens of
Plasmodium falciparum";
RL Mol. Biochem. Parasitol. 27:291-302(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=FCHE5/NF7;
RA Tolle R.;
RN Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX STRAIN=FCHE5/NF7;
RA MEDLINE=93354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
merozoite surface antigen-1";
RL Exp. Parasitol. 81:47-54(1995).
EMBL: Z35328; CAA84557.1; -
InterPro: IPR000561; -
Pfam: PF00008; EGF; 1.
KW Merozoite.

```
FT NON_TER 1 1
SQ SEQUENCE 652 AA; 74292 MW; 2B6A87737B490A62 CRC64;

Query Match 98.08; Score 2011; DB 5; Length 652;
Best Local Similarity 98.2%; Pred. No. 3.9e-100;
Matches 387; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 AISTVMDNLGSGFENEVDVYILKPLAGVYRSLLKQIEKNITFNLDILNSLRKKRY 60
DB 259 AISTVMDNLGSGFENEVDVYILKPLAGVYRSLLKQIEKNITFNLDILNSLRKKRY 318

QY 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 120
DB 319 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 378

QY 121 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 180
DB 379 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 438

QY 181 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
DB 439 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 498

QY 241 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 300
DB 499 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 558

QY 301 HLDRECKCLLNNYKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 360
DB 559 HLDRECKCLLNNYKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 618

QY 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
DB 619 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 652

RESULT 9
Q25969 ID Q25969 PRELIMINARY; PRT; 569 AA.
AC Q25969
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13346; BAA02607.1; -.
DR InterPro: IPR000561; -.
DR Pfam: PF00008; EGF; 1.
DR Merozoite; EGF-like domain.
KW NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 97.58; Score 2001.5; DB 5; Length 569;
Best Local Similarity 98.2%; Pred. No. 1.1e-99;
Matches 387; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 AISTVMDNLGSGFENEVDVYILKPLAGVYRSLLKQIEKNITFNLDILNSLRKKRY 60
DB 177 AISTVMDNLGSGFENEVDVYILKPLAGVYRSLLKQIEKNITFNLDILNSLRKKRY 236

QY 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 120
DB 237 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 296

QY 121 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 180
DB 297 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 355

QY 181 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
DB 356 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
```

```
QY 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 120
DB 237 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 296

QY 121 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 180
DB 297 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 355

QY 181 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
DB 356 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415

QY 241 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 300
DB 416 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEIGKFDQMLNISQHCVKKQCPENSGCFR 475

QY 301 HLDRECKCLLNNYKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 360
DB 476 HLDRECKCLLNNYKQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCECTK 535

QY 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
DB 536 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 569

RESULT 10
Q25974 ID Q25974 PRELIMINARY; PRT; 569 AA.
AC Q25974
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13348; BAA02609.1; -.
DR InterPro: IPR000561; -.
DR Pfam: PF00008; EGF; 1.
DR Merozoite; EGF-like domain.
KW NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 97.58; Score 2001.5; DB 5; Length 569;
Best Local Similarity 98.2%; Pred. No. 1.1e-99;
Matches 387; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 AISTVMDNLGSGFENEVDVYILKPLAGVYRSLLKQIEKNITFNLDILNSLRKKRY 60
DB 177 AISTVMDNLGSGFENEVDVYILKPLAGVYRSLLKQIEKNITFNLDILNSLRKKRY 236

QY 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 120
DB 237 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNLILSKYIKESVENDIKFAQEGIS 296

QY 121 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 180
DB 297 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTPPSPAKTDEQKSKFLPFLTNIETL 355

QY 181 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
DB 356 YNNLVNKIDDDYILNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
```

```

QY 241 LINDTTKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCVKKQCPENSGCFR 300
Db 416 LINDTTKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCVKKQCPENSGCFR 475

QY 301 HLDRECKCLLNKYQBGDKCEENPNTCNENGGCDADATCTEEDSGSRKKITCETCK 360
Db 476 HLDRECKCLLNKYQBGDKCEENPNTCNENGGCDADAKTCTEEDSGSGNGKKTICETCK 535

QY 361 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 569

RESULT 11
Q25975
ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPl.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPl) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED284B2867C9703 CRC64;

```

```

Query Match 97.5%; Score 2001.5; DB 5; Length 569;
Best Local Similarity 98.2%; Pred. No. 1.1e-99;
Matches 387; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 AISVTMDNLSGFENEYDVYILKPLAGVYRSLLKQIEKNIITNLNDILNLSRLKRRY 60
Db 177 AISVTMDNLSGFENEYDVYILKPLAGVYRSLLKQIEKNIITNLNDILNLSRLKRRY 236

QY 61 FLDVLESDLMQFHHISSNEYIIEFSKLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFHHISSNEYIIEFSKLLNSEQNTLLKSKYIKESVENDIKFAQEGIS 296

QY 121 YIEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YIEKVLAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 355

QY 181 YNLVKNIDYILNLKAKINDCNVEKDEAHVKITKLSLKAIDDKIDLKNTNDFEAIKK 240
Db 356 YNLVKNIDYILNLKAKINDCNVEKDEAHVKITKLSLKAIDDKIDLKNTNDFEAIKK 415

QY 241 LINDTTKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCVKKQCPENSGCFR 300
Db 416 LINDTTKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNISOHCVKKQCPENSGCFR 475

QY 301 HLDRECKCLLNKYQBGDKCEENPNTCNENGGCDADATCTEEDSGSRKKITCETCK 360
Db 476 HLDRECKCLLNKYQBGDKCEENPNTCNENGGCDADAKTCTEEDSGSGNGKKTICETCK 535

QY 361 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 569

```

```

RESULT 13
Q25979
ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPl.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

```

OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: D13352; BAA02613.1; -.
DR InterPro: IPR000561; -.
DR Pfam: PF00008; EGF 1.
DR Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match          97.5%; Score 2001.5; DB 5; Length 569;
Best Local Similarity 98.2%; Pred. No. 1.1e-99;
Matches 387; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

Qy 1 AISTVMDNILSGFENEVDVYILKPLAGVYRSLLKQIEKNIITFNLDIILNSRLKRRKY 60
Db 177 AISTVMDNILSGFENEVDVYILKPLAGVYRSLLKQIEKNIITFNLDIILNSRLKRRKY 236
Qy 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNIILKSKYIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSKYIKESVENDIKFAQEGIS 296
Qy 121 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 180
Db 297 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 355
Qy 181 YNLLVNKIDDDYLINLAKKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 356 YNLLVNKIDDDYLINLAKKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 415
Qy 241 LINDDTKMDLKGKLLSTGLVQIFPNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 300
Db 416 LINDDTKMDLKGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 475
Qy 301 HLDRECKCLLNTYKQEGDKCEENPNTCNENGGCDADATCTEEDSGSRKKITCECTK 360
Db 476 HLDRECKCLLNTYKQEGDKCEENPNTCNENGGCDADAKTCTEEDSGSNGKKITCECTK 535
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 569

RESULT 14
Q9T2T5 ID Q9T2T5 PRELIMINARY; PRT; 1694 AA.
AC Q9T2T5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=HNL;
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062348; AAC72884.1; -.
DR InterPro: IPR000561; -.
DR Pfam: PF00008; EGF; 1.
DR Merozoite.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192794 MW; 84CFC0E709F5673B CRC64;

```

```

Query Match          97.5%; Score 2001.5; DB 5; Length 1694;
Best Local Similarity 98.2%; Pred. No. 3.4e-99;
Matches 387; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

Qy 1 AISTVMDNILSGFENEVDVYILKPLAGVYRSLLKQIEKNIITFNLDIILNSRLKRRKY 60
Db 1302 AISTVMDNILSGFENEVDVYILKPLAGVYRSLLKQIEKNIITFNLDIILNSRLKRRKY 1361
Qy 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNIILKSKYIKESVENDIKFAQEGIS 120
Db 1362 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSKYIKESVENDIKFAQEGIS 1421
Qy 121 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 180
Db 1422 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITL 1480
Qy 181 YNLLVNKIDDDYLINLAKKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 1481 YNLLVNKIDDDYLINLAKKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1540
Qy 241 LINDDTKMDLKGKLLSTGLVQIFPNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 300
Db 1541 LINDDTKMDLKGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 1600
Qy 301 HLDRECKCLLNTYKQEGDKCEENPNTCNENGGCDADATCTEEDSGSRKKITCECTK 360
Db 1601 HLDRECKCLLNTYKQEGDKCEENPNTCNENGGCDADAKTCTEEDSGSNGKKITCECTK 1660
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 1661 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 1694

RESULT 15
Q9NHX1 ID Q9NHX1 PRELIMINARY; PRT; 1694 AA.
AC Q9NHX1
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MAJOR MEROZOITE SURFACE ANTIGEN.
GN GPI95.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
RT antigen(gp195)gene of Plasmodium falciparum isolate FCCI/HN.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF218248; AAF27526.1; -.
DR InterPro: IPR000561; -.
DR InterPro: IPR003247; -.
DR Pfam: PF00008; EGF; 1.
DR ProDom: PD001527; -.
DR Merozoite.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192766 MW; B51634A49E0F6728 CRC64;

```

```

Query Match          97.5%; Score 2001.5; DB 5; Length 1694;
Best Local Similarity 98.2%; Pred. No. 3.4e-99;
Matches 387; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

Qy 1 AISTVMDNILSGFENEVDVYILKPLAGVYRSLLKQIEKNIITFNLDIILNSRLKRRKY 60
Db 1302 AISTVMDNILSGFENEVDVYILKPLAGVYRSLLKQIEKNIITFNLDIILNSRLKRRKY 1361
Qy 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNIILKSKYIKESVENDIKFAQEGIS 120
Db 1362 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSKYIKESVENDIKFAQEGIS 1421

```

```
Qy 121 YEKVLAKYKDDLESIRKVIKEEKEKPPSPPTTSPAKTDEQKESKFLPFLTNIETL 180
Db 1422 YEKVLAKYKDDLESIRKVIKEEKE -FPSSPPTTSPAKTDEQKESKFLPFLTNIETL 1480
Qy 181 YNNLVNKIDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 1481 YNNLVNKIDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1540
Qy 241 LINDTKKDMGLKLLSTGLVOIFPNTIISKLIEGKFQDMLNISQHCVKKQCPENSGCFR 300
Db 1541 LINDTKKDMGLKLLSTGLVQNFNTIISKLIEGKFQDMLNISQHCVKKQCPENSGCFR 1600
Qy 301 HLDERECKCLLNYKQBGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 360
Db 1601 HLDERECKCLLNYKQBGDKCEENPNTCNENNGGCDADARCTEEDSGSNGKKITCECTK 1660
Qy 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
Db 1661 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1694
```

Search completed: August 8, 2001, 12:33:44
Job time: 272 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:56 ; Search time 34.41 Seconds
(without alignments)
235.763 Million cell updates/sec

Title: US-09-500-376-3

Perfect score: 2052

Sequence: 1 A1SVTMDNLSGFENEDYVI.....SNFLGIFLLILMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	298	14.5	53	1	US-08-290-919-4
2	282	13.7	53	1	US-08-290-919-3
3	271	13.2	48	1	US-08-290-919-12
4	268	13.1	48	1	US-08-290-919-2
5	265	12.9	48	1	US-08-290-919-1
6	255.5	12.5	106	1	US-08-290-919-11
7	127.5	6.2	740	1	US-08-257-073-5
8	127.5	6.2	2710	1	US-08-480-604A-6
9	127.5	6.2	2710	2	US-08-405-496A-6
10	126.5	6.2	798	2	US-07-728-215-30
11	120.5	5.9	1494	3	US-08-755-587-186
12	118	5.8	1394	4	US-09-213-053-2
13	116	5.7	540	1	US-08-286-325A-2
14	115.5	5.6	677	3	US-08-480-640A-115
15	115.5	5.6	677	3	US-08-480-640A-193
16	115.5	5.6	677	3	US-08-295-802-115
17	115.5	5.6	677	4	US-08-686-968C-58
18	115.5	5.6	677	4	US-08-686-968C-193
19	115.5	5.6	677	4	US-08-488-237A-115
20	115.5	5.6	677	4	US-08-488-237A-193
21	115.5	5.6	984	1	US-08-257-073-3
22	115.5	5.6	984	2	US-08-184-009-120
23	115.5	5.6	984	2	US-08-458-356-120
24	115.5	5.6	1786	4	US-08-973-462-8
25	115.5	5.6	2482	1	US-08-328-254-6
26	115.5	5.6	3248	1	US-08-353-700-1
27	115.5	5.6	3248	5	PCT-US95-16216-1

28	112.5	5.5	912	1	US-07-789-915A-8	Sequence 8, Appli
29	112.5	5.5	912	1	US-08-005-002C-8	Sequence 8, Appli
30	112.5	5.5	912	1	US-08-487-203A-8	Sequence 8, Appli
31	112.5	5.5	1657	1	US-08-287-959-1	Sequence 1, Appli
32	111.5	5.4	1093	4	US-09-315-793-52	Sequence 52, Appli
33	111	5.4	1312	2	US-08-592-126-148	Sequence 148, App
34	111	5.4	1312	2	US-08-687-080-51	Sequence 51, Appli
35	110	5.4	976	4	US-09-104-324B-4	Sequence 4, Appli
36	110	5.4	984	1	US-08-242-932-2	Sequence 2, Appli
37	110	5.4	984	1	US-08-714-481-2	Sequence 2, Appli
38	110	5.4	984	5	PCT-US95-06111-2	Sequence 2, Appli
39	109.5	5.3	921	2	US-08-568-459A-8	Sequence 8, Appli
40	109.5	5.3	921	2	US-08-487-826B-8	Sequence 8, Appli
41	109	5.3	1572	2	US-08-290-731C-5	Sequence 5, Appli
42	109	5.3	1579	3	US-08-755-587-184	Sequence 184, App
43	109	5.3	1596	3	US-09-356-952-3	Sequence 3, Appli
44	108.5	5.3	699	3	US-08-851-843A-52	Sequence 52, Appli
45	108.5	5.3	699	4	US-08-974-549A-189	Sequence 189, App

ALIGNMENTS

RESULT 1
US-08-290-919-4
; Sequence 4, Application US/08290919
; Patent No. 5720959
; GENERAL INFORMATION:
; APPLICANT: HOLDER, ANTHONY A.
; APPLICANT: BLACKMAN, MICHAEL J.
; APPLICANT: CHAPPEL, JONATHAN A.
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
; TITLE OF INVENTION: VACCINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBAY & CUSHMAN, L.L.P.
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,919
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9203821.5
; FILING DATE: 22-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB93/00367
; FILING DATE: 22-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 53 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; FEATURE:
; NAME/KEY: Modified-site

Matches	49;	Conservative	1;	Mismatches	2;	Indels	0;	Gaps	0;
QY	325	PNPTCNENNGCDADATCTEEDSGSSRKKTCTECTKPDSPYPLFDGIFCSSN	376						
Db	2	PNPTCNENNGCDADAKTEEDSGSGNKKTCTECTKPDSPYPLFDGIFCSSN	53						
RESULT	3								
	US-08-290-919-12								
	; Sequence 12, Application US/08290919								
	; Patent No. 5720959								
	; GENERAL INFORMATION:								
	; APPLICANT: HOLDER, ANTHONY A.								
	; APPLICANT: BLACKMAN, MICHAEL J.								
	; APPLICANT: CHAPPEL, JONATHAN A.								
	; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA								
	; TITLE OF INVENTION: VACCINE								
	; NUMBER OF SEQUENCES: 19								
	; CORRESPONDENCE ADDRESS:								
	; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.								
	; STREET: 1100 NEW YORK AVENUE, N.W.								
	; CITY: WASHINGTON								
	; STATE: D.C.								
	; COUNTRY: U.S.A.								
	; ZIP: 20005-3918								
	; COMPUTER READABLE FORM:								
	; MEDIUM TYPE: Floppy disk								
	; COMPUTER: IBM PC compatible								
	; OPERATING SYSTEM: PC-DOS/MS-DOS								
	; SOFTWARE: Patentin Release #1.0, version #1.25								
	; CURRENT APPLICATION DATA:								
	; APPLICATION NUMBER: US/08/290,919								
	; FILING DATE: 04-OCT-1994								
	; CLASSIFICATION: 435								
	; PRIOR APPLICATION DATA:								
	; APPLICATION NUMBER: GB 9203821.5								
	; FILING DATE: 22-FEB-1992								
	; PRIOR APPLICATION DATA:								
	; APPLICATION NUMBER: PCT/GB93/00367								
	; FILING DATE: 22-FEB-1993								
	; ATTORNEY/AGENT INFORMATION:								
	; NAME: KOKOLIS, PAUL N.								
	; REGISTRATION NUMBER: 16,773								
	; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/								
	; TELECOMMUNICATION INFORMATION:								
	; TELEPHONE: (202) 861-3000								
	; TELEFAX: (202) 822-0944								
	; TELEX: 6714627 CUSH								
	; INFORMATION FOR SEQ ID NO: 12:								
	; SEQUENCE CHARACTERISTICS:								
	; LENGTH: 48 amino acids								
	; TYPE: amino acid								
	; STRANDEDNESS: single								
	; TOPOLOGY: linear								
	; MOLECULE TYPE: peptide								
	US-08-290-919-12								
Query Match	13.2%;	Score	271;	DB	1;	Length	48;		
Best Local Similarity	95.8%;	Pred. No.	7.3e-14;						
Matches	46;	Conservative	1;	Mismatches	1;	Indels	0;	Gaps	0;
QY	281	NISOHCVKKQCPNSGCFRHLDERECKCLLNKYQEGDKCEENPNT	328						
Db	1	NISOHCVKKQCPNSGCFRHLDERECKCLLNKYQEGDKCEENPNT	48						
RESULT	4								
	US-08-290-919-2								
	; Sequence 2, Application US/08290919								
	; Patent No. 5720959								
	; GENERAL INFORMATION:								
	; APPLICANT: HOLDER, ANTHONY A.								

```

CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
APPLICATION DATA:
APPLICATION NUMBER: PC7/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= X
OTHER INFORMATION: /note= "X" = M and N, or N"
US-08-290-919-1

Query Match 12.9%; Score 265; DB 1; Length 48;
Best Local Similarity 95.7%; Pred. No. 2e-13;
Matches 45; Conservative 1; Mismatches 1; Indels

QY 282 ISOHCVKKQCPNSGCFHLDERECKLLNYKQEGDKCENPNT 328
|||||
DB 2 ISOHCVKKQCPNSGCFHLDERECKLLNYKQEGDKCENPNT 48
|||||

RESULT 6
US-08-290-919-11
Sequence 11, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

Query Match	6.2%;	Score 127.5;	DB 1;	Length 740;
Best Local Similarity	21.8%;	Pred. No. 0.1;		

Query Match 6.2%; Score 127.5; DB 1; Length 740;
Best Local Similarity 21.8%; Pred. No. 0.1;
Matches 95; Conservative 58; Mismatches 114; Indels 169; Gaps 22;

[illegible]

QY	204	V-----	-EKDEAHVKITKLS-DLKAIDDK	225
Db	517	VVPPTOSKKKNKNETYSGMDENFHNHPENYFKREYYIDENDMDMEVKVKIGVTLK----	K	572
QY	226	IDLPKNTNDPEALK-----	KLINDDPK-----	262
Db	573	FEPKNGNVSETKLHLGNKKKKHTEAINDIITQKELQAIYNELMNYTNGCNKNIQOI	6322	
QY	263	FPNTIISKLEGFQDMLNITSQHCVKKQ-----	CPENSGCFRHLDEREE	307
Db	633	FOQNILE-----	NDVLNQETEEMEKEQVEAITKQIEAVDALAPKNK-----	679
QY	308	CKCLLNYKQEGDKCEE	323	
Db	680	EKEKEKEKEKEKEEK	695	
RESULT 8				
US-08-480-604A-6				
; Sequence 6, Application US/08480604A				
; Patent No. 5736139				
; GENERAL INFORMATION:				
; APPLICANT: KINK, JOHN A.				
; APPLICANT: THALLEY, BRUCE S.				
; APPLICANT: PADHYE, NISHA V.				
; APPLICANT: FIRCA, JOSEPH R.				
; APPLICANT: STAFFORD, DOUGLAS C.				
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND				
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE				
; NUMBER OF SEQUENCES: 32				

```

; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-604A-6

Query Match 6.2%; Score 127.5; DB 1; Length 2710;
Best Local Similarity 23.4%; Pred. No. 0.52;
Matches 74; Conservative 47; Mismatches 102; Indels 93; Gaps 17;
QY 36 IEKNIITFNLDI-INSRLKKRKYFL-----DVLESDLMQFKHISNEYIIE 84
DB 732 VKNKSITIGANQYEVIRNS--EGRKELLAHSGKWINEEATMSDL-----SSKEYIFFD 783
QY 85 SF-KLLANSEOKNI--LLKSKYKIK-----ESVENDIKFAQEGIS-----YVEK- 124
DB 784 SIDNKLAKSKNIPGLASISEDIKTLILDASVSPDTKFIPLNKLNTIESSIGDYIYVEKL 843
QY 125 -----VLAKYKDDLESIKKVIKEKEKFPSPPTTPPSA----- 159
DB 844 EPVKNITHNSIDLDIFENLLENVSDLEYELKLNNDKYLISFEDISKNNSTYSVRFI 903
QY 160 -KTD-----EQRKE--SKFLPFLT-NIETLYNNLVNKIDDIYLNKAKINDCNVERDEA 209
DB 904 NKSNGESVYVTEKEIFSKYSEHITKESTIKNSIITDVNGNLLD-----NIQLDHT 955
QY 210 HVKTKISDLKAIODKIDLFKNWDFRAKKLIND---DTKMKMLGKLLSTGLVQIOPNT 266
DB 956 S-QVNTLNAAFIQLSDYSSN-----KDVLDLSTSVKQVLYAQLFSTGLNTIYDSI 1007

QY 267 IISKIEGKFODMLNI 282
DB 1008 QLVNLSNAVNDTINV 1023
RESULT 9
US-08-405-496A-6
Sequence 6, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-405-496A-6

Query Match 6.2%; Score 127.5; DB 2; Length 2710;
Best Local Similarity 23.4%; Pred. No. 0.52;
Matches 74; Conservative 47; Mismatches 102; Indels 93; Gaps 17;
QY 36 IEKNIITFNLDI-INSRLKKRKYFL-----DVLESDLMQFKHISNEYIIE 84
DB 732 VKNKSITIGANQYEVIRNS--EGRKELLAHSGKWINEEATMSDL-----SSKEYIFFD 783
QY 85 SF-KLLANSEOKNI--LLKSKYKIK-----ESVENDIKFAQEGIS-----YVEK- 124
DB 784 SIDNKLAKSKNIPGLASISEDIKTLILDASVSPDTKFIPLNKLNTIESSIGDYIYVEKL 843
QY 125 -----VLAKYKDDLESIKKVIKEKEKFPSPPTTPPSA----- 159

Db 844 EPVANIITHNSIDDLIDFENLLENVSDYELKLLNNLDEKYLISFEDISKNNSTYSVRFI 903
QY 160 -KTD-----EOKKE--SKLPFLT-NIETLYNNLVNKIDYLLINKAKINDCNVKEDEA 209
Db 904 NKSNGSVYVETEKEIEFSKSEHITKEISTIKNSIIDVNGNLLD-----NIQDHT 955
QY 210 HVKITKLSDLKAIDDKIDLPKNTNDFEAIKKLIND---DTRKMDLGLKLLSTGLVQIPNPNT 266
Db 956 S-QVNTLNAFFIQLSDIYSSN-----KDVLDLSTSVKVLQYLAQLFSTGLNTIYDSI 1007
QY 267 IISKLEGGFQDMLNI 282
Db 1008 QLVNLISNAVNDITNV 1023
RESULT 10
US-07-728-215-30
; Sequence 30, Application US/07728215
; Patent No. 5962643
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Dean
; APPLICANT: Quaranta, Vito
; APPLICANT: Pytela, Robert
; TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States of America
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07728,215
; FILING DATE: 19910711
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P31 8717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-728-215-30
Query Match 6.2%; Score 126.5; DB 2; Length 798;
Best Local Similarity 21.0%; Pred. No. 0.13;
Matches 105; Conservative 69; Mismatches 137; Indels 189; Gaps 32;
QY 48 NDILNSR----LKKRKVFLDVLSDLMQFK-----HTSSNEYIEDSFKLINSEOKNILL 98
Db 78 DDIENTPGSKDIKNKNVTRSGTAELKLPEDIHQIPOOLVL-----RLRSGEPQFTL 133
QY 99 KSKYIKESVENDIKFAQGISYKVKLAKYKDDLESIKV---IKEEKKEKFPSS----- 150
Db 134 KFRR--AEDYPIDLYLMD-LSY-----SMKDDLENVKSIGTDLNMEMRRITSDFRIGF 184
QY 151 -----PPTPPSPAK-----TDEQKESKF-----LPETLNIEIYNNLVNK----- 187
Db 185 GSFEKVTMPYIISTTPAKLRNPCTSEQNCTTPFSYKNVLSTLNKGEVFNELVGKQRISGN 244

QY 188 ID-----DYLNL-----KAKI-----ND--CNVE 205
Db 245 LDSPEGFDALMOVAVCGSLIGHNRVTRLLVFTSDAGFHAGDKLGGIVLPNDGQCHLE 304
QY 206 KDE-----AHVKITKLSDLKAIDDKIDLPKNTNDFEAIKKLINDDTKKMDLKG 253
Db 305 NMYTWSHYDYPSIAHL-VQKLSE-----NNIQTITAVTEEPQVYKELKNLIPKSAVGT 359
QY 254 LL--STGLVQIF--PNTIISKLI--EGKFQDMLNISOHOCVK----- 289
Db 360 LSANSSNVLIQIIDAYNSLSLEVLNGLSEGVTSYKSYCKNGVNGTGENGRKCSNIS 419
QY 290 -----KQCP-ENSGCFR-----HLDERE-----ECKLLNYKQEGDKCEE 323
Db 420 IGDEVQFEISITSNKCPKSDSKIRPLGFTVEEVILQYICECEGQSGEIPSPACHE 479
QY 324 NPNT-----CNENNGG---C-----DADATCTEEDSG---SSRKKIT---CEC 358
Db 480 G-NGTECGACRCNEGRVGRHCECSTDEVNSEMDMDAYCRKENSEISCSNNGECVCGQVC 538
QY 359 TKPD-SYPLFDGIFCSSNF 377
Db 539 RKRDNTEIYSGKEFCEDNF 558
RESULT 11
US-08-755-587-186
; Sequence 186, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A.
; APPLICANT: Wooster, Richard F.
; APPLICANT: Ashworth, Alan
; APPLICANT: Stratton, Michael R.
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 186:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1494 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-755-587-186

Query Match 5.9%; Score 120.5; DB 3; Length 1494;
Best Local Similarity 22.6%; Pred. No. 0.81;
Matches 94; Conservative 68; Mismatches 167; Indels 87; Gaps 21;

Qy 4 VTMDNLGFM-----EYDVYIKPLAGVYRSUKKOIERNIITFNLN-----DILNSR 54
Db 760 VTMSKQOLTANTGQNIKIDFTFYL-----SFQTASRKNIYRSLSNKAISLLNQK 811

Qy 55 LKRY--FLDVLSDLMQFKHIS---SNEVLIEDSKLLNSEOK-----NI 96
Db 812 WTEELNPFNSLSNELLPGDIDKTDISNHEVIE-----NTERKDKITKESLTGTENI 865

Qy 97 LKSYKIKESVENDIKFAQEG--ISYVE--KVLAKYKDDLESIKVKIKEKEKFPSSP 151
Db 866 LL---ILQRPESKIKIKESAVLGFTASGKKTEIKESLDKVKLNF-EKEQDNSEI 920

Qy 152 PTPPSPAKTDEQKESK-----FLPFLNIETLYNNL-VNKDDYDYLINLAKIN 200
Db 921 TNFSHRGAKMSKDRCEKDGRELACGTTEITTPYEETHSSLEKKKVYSNEIAALRPLL 980

Qy 201 DGNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKLLINDDTKKDMLGKLLST-GL 259
Db 981 SDNLYKQTEENLAKISHA-----SQKVDVHENTE-----KETAKPTMTNOSTYSA 1026

Qy 260 VOIFPNTIISLIEGKFDQMLNISQHCQKCPENSGCFRHLDERECKICLLNYKQEGD 319
Db 1027 IENSPLT-FTQDTEKF---SVSEASLFEAKKWLREGEWDDQSERINA-AKVNCKEYPD 1080

Qy 320 KCEPNPTCNENGGCDAATCTEEDSGSRKKTCTCTKPDSPYPLFDGIPCSSS 375
Db 1081 DYVENPSCGNSNSAITENDKHLSEKOGST---YLSNSTMSNSYSYHPG-FCHSS 1132

RESULT 12
US-09-213-053-2
; Sequence 2, Application US/09213053
; Patent No. 6159477
; GENERAL INFORMATION:
; APPLICANT: AUDONNET, Jean-Christophe
; APPLICANT: BAUDU, Philippe
; TITLE OF INVENTION: CANINE HERPESVIRUS BASED RECOMBINANT LIVE VACCINE, IN
; TITLE OF INVENTION: PARTICULAR AGAINST CANINE DISTEMPER, RABIES OR THE
; TITLE OF INVENTION: PARAINFLUENZA 2 VIRUS
; FILE REFERENCE: 454313-2200
; CURRENT APPLICATION NUMBER: US/09/213.053
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 9608242
; EARLIER FILING DATE: 1996-06-27
; EARLIER APPLICATION NUMBER: PCT/FR97/01115
; EARLIER FILING DATE: 1997-06-23
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1394
; TYPE: PRT
; ORGANISM: Canine herpesvirus
US-09-213-053-2

Query Match 5.8%; Score 118; DB 4; Length 1394;
Best Local Similarity 19.2%; Pred. No. 1.1;
Matches 102; Conservative 65; Mismatches 153; Indels 212; Gaps 25;

Qy 18 DVIYKPLAGVYRSUK-----KQTEKNIIITFNMLNDILNSRLKRRYF--LDV-- 64
Db 866 DTTIKPISTEANNLKSYSTSIKPKPLKLLKSKCRDNVIY-----RPYFSQLEINC 919

Qy 65 -----LESQLMQ--FHHISSNEVLIEDSKLLNSEOKNILLKSYKIKESVEND---- 111
Db 920 TITKKNLSNPLIELWFKELTKYNTKNENVESLKTDTDISKNTNSDNFYNDLLG 979

Qy 112 -----IKFAQEGISYVE--KVLAK-----YKDDLESIKKVIKEKEKFP 148

Db 980 IQNPVNTKLYGSQFYDNGNILLNIKSVDKFTSGTYTWKLYNNSNESIFETFKIQVAYH 1039
Qy 149 SSPPTTPSPAKTDBQKE-----SKFLPFLNIETLYNNLVNKIDDIYLINLAKINDCNV 204
Db 1040 SPNVNLKSNPNSLYNYSIACTIANIYFP-LESTEIWFNDGQPIDKKYI----- 1087
Qy 205 EKDEAH-----VKITKLSDLKAIDD-----KIDLFKNTNDFEAIKLLINDDTKK 248
Db 1088 --DETYSWIDGLITRTSILSPSEAMESPPNLCNVETWYKNS---KASKK----- 1134
Qy 249 DMLGKLLSTGLVQIIPPNTIISKL-----IEGQFQMLNISQHCQVK-----KQCP 293
Db 1135 -----FSNTVIPKVVYKPFISIKFDNGLAICDAKCVSRNNKQLWLKVDIP 1180
Qy 294 EN-----SG-CFRH-----LDERECKCLL-----NYKQEG--DKCEE 323
Db 1181 INGGDIISGCLNHPGLVNIQNIKIDISYDEFTVYKCSIIIGYPIIFPFYDEKVFDSDE 1240
Qy 324 NPNPT-----CNENNGGC-----DADATC 342
Db 1241 NVSKSMLISITIIIGGAIFVIVIFITALCFYCSKNNKIMAOQLVLTIDPLEDVENKNTSS 1300
Qy 343 TEEDSGSRKKTCTCTKPDSPYPLFDGIFCSSNFISFLILMLILYSFI 394
Db 1301 DEETNLNOKKSTQC-----LCVTLGFFAAGILLITIAAIIFTFI 1340

RESULT 13
US-08-286-325A-2
; Sequence 2, Application US/08286325A
; Patent No. 5658770
; GENERAL INFORMATION:
; APPLICANT: PREVOTS, Fabien
; APPLICANT: REMY, Elisabeth
; APPLICANT: RITZENTHALER, Paul
; TITLE OF INVENTION: Nucleic acid sequence and plasmids
; TITLE OF INVENTION: comprising at least one phage resistance mechanism,
; TITLE OF INVENTION: bacteria in which they are present, and their use
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BACON & THOMAS
; STREET: 625 Slaters Lane - Fourth Floor
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286.325A
; FILING DATE: 04-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 93 09 777
; FILING DATE: 09-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: FICHTER, Richard E
; REGISTRATION NUMBER: 26,382
; REFERENCE/DOCKET NUMBER: REF/BDL/
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 683-1080
; TELEFAX: (703) 683-1080
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 540 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-325A-2

```
;
; POSITION IN GENOME:
; MAP POSITION: -23.2
; UNITS: %G
; US-08-480-640A-115

Query Match      5.6%  Score 115.5; DB 3; Length 677;
Best Local Similarity 21.6%  Pred. No. 0.69;
Matches 71; Conservative 59; Mismatches 123; Indels 75; Gaps 16;

QY 22 LKPLAGVYRSKKQIEKNITFNILNLDILNSRLKRRKYFLDVLVLE---SDLMQPKHSSN 78
Db 199 LRDINRIIEELKKYPNNII-----DYISDSIKSSSFHILHMIISNMFPAIIPSVN 251
QY 79 EYI--IEDSPKLLNSEQ-KNILLKASYK-----IKESVENDIKFAQEGISYEKVLAKYKD 131
Db 252 DFISTVVDKDLINMYGIKCVAMFSYDINMIDLESDDSDYIFIEKNISYD---VKCRD 308
QY 132 DLESIKVKIEKEKEKFPSPPTTPPSAKTDEQKESKFLPFLTNIETLYNNLVNKIDDY 191
Db 309 FANMIRKVKREKNRIIT-----KCEDIIRYIKLES-----KNRINDENKNKVEEV 354
QY 192 LINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAKKLLINDDTKKDML 251
Db 355 LIHID-----NVSKNNK-----LSLSDISSLMQDFRL-----NPCTIRNILLSSATIK--- 397
QY 252 GKLLSTGLVQ-----IFPNTIISKIEGKF-QDMLNISQHCQVK--KOCPENSCGFRHLD 303
Db 398 SKLLALRAVNKNWKCYSLTNVSMYKRIKGVIMDMVDYISTNLIKHYKOLYDKMSTFEYKR 457
QY 304 ERECKC-----LLNYK 315
Db 458 DIKSKCSICSDSITHIYETTSCINVK 485

RESULT 15
US-08-480-640A-193
; Sequence 193, Application US/08480640A
; Patent No. 6033904
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,640A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-640A-193

Query Match      5.7%  Score 116; DB 1; Length 540;
Best Local Similarity 25.0%  pred. No. 0.48; 89; Indels 76; Gaps 18;
Matches 70; Conservative 45; Mismatches 89; Indels 76; Gaps 18;

QY 28 VYRSKKQIEKN---IITFNILNLDILNSRLKRRKYFLDVLVLESDLMQPKHSSNEYIED 84
Db 250 VYHICKKLMKKNKRAIVHGANLNN-GQRLALRGFTTFPVKSIIEVLNADKYDIYVD 308
QY 85 SFKLLNSQKILLKSYKIKESVEN-DIKF--AQEGISYEKVLAKYKDDLESTK---KV 139
Db 309 EAORLRD-----LGRQYTKLVDTTENSOTKEFIISLDG---RQTLNKYEIEENSIKLIFY 360
QY 140 IKEE-----KEKPPSSPTTPPSAKTDEQKESKFLPFLTNIETLYN--NLVKNKIDDY 191
Db 361 INKGVTSLKDKRTNP-----EMSKFIQLLFKI--PMYKKIDLISNIDH- 404
QY 192 LINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAKKLLINDDTKKDML 251
Db 405 --NIIKYFD-NRESGNEVI-----SDM-----DSNSDNEVL-NYTKDRFRKTGI 445
QY 252 GKLLSTGLV-----QIFPNTIIL-----SKLIEGK 275
Db 446 GKMGCGNLTSHITGOEPDKVILPLDSNFFYKEQKIIDS 485

RESULT 14
US-08-480-640A-115
; Sequence 115, Application US/08480640A
; Patent No. 6033904
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,640A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; CLONE: 515-85.1
```


Search completed: August 8, 2001, 12:29:58
Job time: 46 sec

THIS PAGE BLANK (0870)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:31:53 ; Search time 60.73 seconds
(without alignments)
393.312 Million cell updates/sec

Title: US-09-500-376-3

Perfect score: 2052

Sequence: 1 A1SVTMDNLTSGFENEYDVI.....SNFLGTSFLLMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601:*

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1248.5	60.8	1639	AAW54145	P. falciparum synt
2	1156.5	56.4	355	AAV09372	Merozoite surface
3	1156.5	56.4	355	AAV05832	Merozoite surface
4	1156.5	56.4	361	AAV09373	Merozoite surface
5	1156.5	56.4	361	AAV05833	Merozoite surface
6	1149.5	56.0	376	AAV09374	Modified merozoite
7	1149.5	56.0	376	AAV05834	Modified merozoite
8	1091	53.2	1654	AAV50777	Sequence of the PI
9	641.5	31.3	282	AAW37610	Merozoite surface
10	615	30.0	116	AAW36103	PfMSP1(p19)A prote
11	615	30.0	116	AAW22592	PfMSP1(p19)A prote

12	529	25.8	108	22	AAW37609	Merozoite surface
13	528	25.7	96	22	AAW37608	Merozoite surface
14	521	25.4	127	18	AAW22593	PfMSP1(p19)S prote
15	521	25.4	127	18	AAW36102	PfMSP1(p19)S prote
16	304	14.8	54	14	AAW41357	MSP1EGF2B EGF2-11k
17	288	14.0	54	14	AAW41356	MSP1EGF2A EGF2-11k
18	276	13.5	49	14	AAW41355	MSP1EGF1B EGF1-11k
19	273	13.3	49	14	AAW41354	MSP1EGF1A EGF1-11k
20	271	13.2	93	22	AAW37611	Merozoite surface
21	255.5	12.5	106	14	AAW41358	P. yoelli combined
22	187.5	9.1	350	21	AAW70278	Recombinant vaccin
23	166	8.1	980	21	AAW18294	Plasmodium falcipa
24	145.5	7.1	1712	21	AAW18205	Plasmodium falcipa
25	139	6.8	1979	21	AAW18171	Plasmodium falcipa
26	134	6.5	2573	21	AAW18234	Plasmodium falcipa
27	134	6.5	4134	20	AAW31946	Plasmodium falcipa
28	132.5	6.5	1188	21	AAW18183	Plasmodium falcipa
29	132.5	6.5	1346	21	AAW18236	Plasmodium falcipa
30	132	6.4	1166	21	AAW18268	Plasmodium falcipa
31	129.5	6.3	1817	21	AAW18255	Plasmodium falcipa
32	128.5	6.3	646	21	AAW18188	Plasmodium falcipa
33	128.5	6.3	2539	21	AAW18198	Plasmodium falcipa
34	127.5	6.2	740	13	AAW27530	Plasmodium falcipa
35	127.5	6.2	740	16	AAW68838	Plasmodium falcipa
36	127.5	6.2	2710	17	AAW95016	Plasmodium falcipa
37	127.5	6.2	2710	19	AAW68387	C. difficile toxin
38	127	6.2	26	22	AAW71974	Clostridium diffic
39	127	6.2	2500	21	AAW18272	Plasmodium falcipa
40	126.5	6.2	798	22	AAW36937	Plasmodium falcipa
41	126.5	6.2	963	19	AAW70540	Human integrin bet
42	126.5	6.2	1743	19	AAW98879	Integrin beta-1 ch
43	126	6.1	2285	20	AAW98149	H. pylori GHPO 175
44	124.5	6.1	389	19	AAW98813	Bacillus subtilis
45	123.5	6.0	717	21	AAW21231	H. pylori GHPO 137
						Tomato LAMPFL. Ly

ALIGNMENTS

RESULT 1	
AAW54145	AAW54145 standard; Protein; 1639 AA.
ID	
AC	AAW54145;
XX	
DT	23-SEP-1998 (first entry)
XX	
DE	P. falciparum synthetic gp190 protein.
XX	
KW	gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
KW	monoclonal antibody; passive immunisation; parasite.
XX	
OS	Plasmodium falciparum.
OS	Synthetic.
PN	WO9814583-A2.
PN	
XX	
PD	09-APR-1998.
XX	
PF	02-OCT-1997; 97WO-EP05441.
XX	
PR	02-OCT-1996; 96DE-4040817.
XX	
PA	(BUJA/) BUJARD H.
XX	
PI	Bujard H, Pan W, Tolle R;
XX	
DR	WPI: 1998-240088/21.
DR	N-PSDB; AAV21451, AAV35363.
XX	
PT	Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT	protein - useful in anti-malaria vaccines, also stabilising genes by
PT	reducing their AT content

XX Example 1; Fig 3c; 48pp; German.
 XX This sequence represents a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface) protein. The gene encoding this protein has been
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 CC protein is useful in vaccines against malaria or for producing monoclonal
 CC antibodies (for passive immunisation). The complete gp190 protein can now
 CC be produced outside the parasite and has, at least over extended regions,
 CC the native pattern of folding. Larger amounts of the protein can be
 CC produced recombinantly than would be possible using the parasites as
 CC source.
 XX Sequence 1639 AA;
 SQ

Query Match 60.8%; Score 1248.5; DB 19; Length 1639;
 Best Local Similarity 61.0%; Pred. No. 3e-79;
 Matches 239; Conservative 54; Mismatches 74; Indels 25; Gaps 3;

QY 6 MDNLGSGFENEYDVIYKPLAGVYRSLLKQIEKNIIITFNLDNLNLSRLKRYFLDVL 65
 Db 1270 idnlksieneyevlylkplagvyrslkqlennvmfnnvkdilnsrfnkrenfknvl 1329

QY 66 ESDLMPQKHSSNEYIIEEDSKLLNSEQKLLKSYKIKESVENDIKFAQEGISYVEKV 125
 Db 1330 esdlipykdlitssnyvvkdpkyflnkekrdfkflssynyikdsidtdinfandvlygkyl 1389

QY 126 LAKYKDDLESIKKVIKEKEKFPSSPTTPPSAKTDEQKESKFLPFLTNIETLYNNLV 195
 Db 1390 sekyksdlsdkkyi-----ndkqgenekylpflnnietlyktvn 1429

QY 186 NKIDYDLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNITNDFEATKLLINDD 245
 Db 1430 dkidlfvihleakvlnyteknsvevkikelnlyktldkfknfnfvgiadlstdy 1489

QY 246 TKDMLGKLLSTGLVQIFPN---TIISKLEGRFQDMLNISQHCVKKQCPENSGCFRHL 302
 Db 1490 nhnlltkflstgm---fenlaktvlsnldgnlqgmnlisqhcvkqkqcpnsgcfrhl 1547

QY 303 DERECKLLNYKQEGDKCEENPNTCNENNGCDADATCTEEDSGSRKKITCECTKPD 362
 Db 1548 dereeckllnykqegdkcvenpnptcnennggcdadadakcteedsgngkkitcectkpd 1607

QY 363 SYPLFDGIFCSSNFGISFLIIMLILYSFI 394
 Db 1608 syplfdgifcssnfgliffllimlilysfi 1639

RESULT 2
 AAY09372
 ID AAY09372 standard; Protein; 355 AA.
 XX
 AC AAY09372;
 XX
 XX 31-AUG-1999 (first entry)
 XX
 XX Merozoite surface protein MSP-1-42.
 XX
 KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal.
 XX
 XX Plasmodium falciparum.
 XX
 XX WO9920774-A2.
 XX
 XX 29-APR-1999.
 XX
 XX 20-OCT-1998; 98WO-US222226.
 XX
 XX 15-MAY-1998; 98US-0085649.
 PR
 XX 20-OCT-1997; 97US-0062592.

XX (GENZ) GENZYME TRANSGENICS CORP.
 XX Chen LH, Meade H;
 PI WPI; 1999-288313/24.
 DR P-PSDB; AAX56008.
 XX Modified malarial protein for use in anti-malarial vaccines
 PT Example; Fig 1; 35pp; English.
 PS
 XX The present sequence represents a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
 CC has been modified (see AAX56008) compared to the native sequence (see
 CC AAX56009) such that 306 nucleotide positions have been replaced to
 CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
 CC instability motifs while maintaining the same protein amino acid
 CC sequence. These alterations allow MSP-1-42 to be expressed in
 CC mammalian cell culture and in transgenic mice. Native MSP-1-12
 CC is known to be difficult to express in cell culture systems,
 CC mammalian cell culture systems or in transgenic animals. The
 CC invention allows expression of MSP-1 protein in the milk of
 CC transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 sequence.
 XX Sequence 355 AA;
 SQ

Query Match 56.4%; Score 1156.5; DB 20; Length 355;
 Best Local Similarity 59.0%; Pred. No. 1.2e-73;
 Matches 219; Conservative 54; Mismatches 73; Indels 25; Gaps 3;

QY 6 MDNLGSGFENEYDVIYKPLAGVYRSLLKQIEKNIIITFNLDNLNLSRLKRYFLDVL 65
 Db 7 idnlksieneyevlylkplagvyrslkqlennvmfnnvkdilnsrfnkrenfknvl 66

QY 66 ESDLMPQKHSSNEYIIEEDSKLLNSEQKLLKSYKIKESVENDIKFAQEGISYVEKV 125
 Db 67 esdlipykdlitssnyvvkdpkyflnkekrdfkflssynyikdsidtdinfandvlygkyl 126

QY 126 LAKYKDDLESIKKVIKEKEKFPSSPTTPPSAKTDEQKESKFLPFLTNIETLYNNLV 185
 Db 127 sekyksdlsdkkyi-----ndkqgenekylpflnnietlyktvn 166

QY 186 NKIDYDLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNITNDFEATKLLINDD 245
 Db 167 dkidlfvihleakvlnyteknsvevkikelnlyktldkfknfnfvgiadlstdy 226

QY 246 TKDMLGKLLSTGLVQIFPN---TIISKLEGRFQDMLNISQHCVKKQCPENSGCFRHL 302
 Db 227 nhnlltkflstgm---fenlaktvlsnldgnlqgmnlisqhcvkqkqcpnsgcfrhl 284

QY 303 DERECKLLNYKQEGDKCEENPNTCNENNGCDADATCTEEDSGSRKKITCECTKPD 362
 Db 285 dereeckllnykqegdkcvenpnptcnennggcdadadakcteedsgngkkitcectkpd 344

QY 363 SYPLFDGIFCS 373
 Db 345 syplfdgifcs 355

RESULT 3
 AAY05832
 ID AAY05832 standard; Protein; 355 AA.
 XX
 AC AAY05832;
 XX
 XX 02-AUG-1999 (first entry)
 DT
 XX Merozoite surface protein MSP-1-42.
 DE

QY 66 ESDLMOFKHSSNEYIIEDSFKLLNSQKWLKSYKIKESVENDIKFAQEGISYKEV 125
 DB 67 esdlipykdlitssnyvkvdkpflnkekrdkflssynyikdsidtdinfandvlgrykll 126
 QY 126 LAKYKDDLESIKKVIKEEKFPSPPTTPPSAKTDEQKESKFLPFLNIETLYNNLV 185
 DB 127 sekyksdlsdskkyl-----ndkqgenekylpflnnietlyktvn 166
 QY 186 NKIDVYLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKLLINDD 245
 DB 167 dkidlfvihleakvlnlyteksnvevkikelnlylktlqkladfkknntfvgladlsty 226
 QY 246 TKDMLGKLLSTGLVQIFPN---TIISKLEGKFQDMLNISQHCVKQCPENSGCFRHL 302
 DB 227 nhnlltkflstgm--fenlaktvlnlsgnlgmlnisghqcvkkgcpqsgcfrhl 284
 QY 303 DERECKCLLYKQEGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTKPD 362
 DB 285 dereeckcllnykgqgdckvenpntcnennngcdadakteedsgsgngkkitcectkpd 344
 QY 363 SYPLFDGIFCS 373
 DB 345 syplfdgifcs 355
 RESULT 5
 AAY05833
 ID AAY05833 standard; Protein; 361 AA.
 AC AAY05833;
 DT 02-AUG-1999 (first entry)
 DE Merozoite surface protein MSP-1-42.
 XX MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal.
 XX Plasmodium falciparum.
 OS Plasmodium falciparum.
 PN WO9920766-A2.
 XX 29-APR-1999.
 XX 20-OCT-1998; 98WO-US22225.
 XX 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.
 XX (GENZ) GENZYME TRANSGENICS CORP.
 PA Chen LH, Meade H;
 PI WPI; 1999-302742/25.
 DR N-PSDB; AAX25587.
 XX New modified recombinant nucleic acid sequences useful for producing
 PT malarial DNA vaccine
 XX Disclosure; Fig 2; 43pp; English.
 PS This present sequence comprises a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The C-terminal end of the sequence is
 CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
 CC AAX25593) encoding MSP-1-42 have been modified according to a method
 CC of the invention in order to improve expression in mammalian cells
 CC and in transgenic animals. The invention provides modified
 CC recombinant nucleic acid sequences and methods for increasing the
 CC mRNA levels and protein expression of proteins that are difficult
 CC to express in cell culture systems, especially mammalian cell

CC culture systems or in transgenic animals. The preferred difficult
 CC protein candidates for expression are those derived from lower
 CC organisms such as parasites, bacteria and viruses that have DNA
 CC coding sequences of high AT content or which have mRNA instability
 CC motifs or rare codons relative to the recombinant expression system
 CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.
 XX Sequence 361 AA;
 SQ
 Query Match 56.4%; Score 1156.5; DB 20; Length 361;
 Best Local Similarity 59.0%; Pred. No. 1.2e-73;
 Matches 219; Conservative 54; Mismatches 73; Indels 25; Gaps 3;
 QY 6 MDNILSGFENEYDVIVILKPLAGVYRSLLKQIEKNIITFNILNLDILNSRLKKRYFLDVL 65
 DB 7 idnilskieneyevlykplagvyrsllkqlennvmtfnvkvdiinsrfnkrenfknl 66
 QY 66 ESDLMOFKHSSNEYIIEDSFKLLNSQKWLKSYKIKESVENDIKFAQEGISYKEV 125
 DB 67 esdlipykdlitssnyvkvdkpflnkekrdkflssynyikdsidtdinfandvlgrykll 126
 QY 126 LAKYKDDLESIKKVIKEEKFPSPPTTPPSAKTDEQKESKFLPFLNIETLYNNLV 185
 DB 127 sekyksdlsdskkyl-----ndkqgenekylpflnnietlyktvn 166
 QY 186 NKIDVYLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKLLINDD 245
 DB 167 dkidlfvihleakvlnlyteksnvevkikelnlylktlqkladfkknntfvgladlsty 226
 QY 246 TKDMLGKLLSTGLVQIFPN---TIISKLEGKFQDMLNISQHCVKQCPENSGCFRHL 302
 DB 227 nhnlltkflstgm--fenlaktvlnlsgnlgmlnisghqcvkkgcpqsgcfrhl 284
 QY 303 DERECKCLLYKQEGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTKPD 362
 DB 285 dereeckcllnykgqgdckvenpntcnennngcdadakteedsgsgngkkitcectkpd 344
 QY 363 SYPLFDGIFCS 373
 DB 345 syplfdgifcs 355
 RESULT 6
 AAY09374
 ID AAY09374 standard; Protein; 376 AA.
 AC AAY09374;
 DT 31-AUG-1999 (first entry)
 DE Modified merozoite surface protein MSP-1-42.
 XX MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.
 XX Plasmodium falciparum.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH Peptide 1..15
 FT /note= "beta-casein signal peptide"
 FT Peptide 371..376
 FT /note= "6xHis tag"
 FT Misc-difference 197
 FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
 FT Misc-difference 278
 FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
 XX WO9920774-A2.
 PN

```
XX 29-APR-1999.
PD KW
PF PE 20-OCT-1998; 98WO-US22226.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-288313/24.
DR P-PSDB; AAX56008.
XX
PT Modified malarial protein for use in anti-malarial vaccines
XX
PS Example; Fig 11; 35pp; English.
XX
CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The sequence has been modified to include
CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC tag. In addition, N181Q and N262Q mutations have been introduced to
CC eliminate 2 N-glycosylation sites. These modifications allow the
CC MSP-1-42 protein to be expressed in the milk of transgenic mice.
XX
SQ Sequence 376 AA;

Query Match 56.0%; Score 1149.5; DB 20; Length 376;
Best Local Similarity 58.8%; Pred. No. 4e-73;
Matches 218; Conservative 53; Mismatches 75; Indels 25; Gaps 3;

Qy 6 MDNILSGFENEYDVYLYKPLAGVYRSKQIEKNIITFNLMNDILNSRLKRRYFLDLV 65
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
22 idnllskieneyevilykplagvyrslkqlennvmtfnvkvdklnsrfrnkrenfknl 81

Qy 66 ESDLMOQKHISSENYEYIIEDESKLLNSEQKNILLKSKYKIKESVENDIKFAQEGISYEKV 125
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
82 esdlipykdltsnyvkvdkpykflnkekrkflsynyikdsidtdinfandvgygkil 141

Qy 126 LAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKTDBQKESKFLPLTNIETLYNNLV 185
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
142 sekyksdidsikkyi-----ndkqgenekyflplnnietlyktn 181

Qy 186 NKIDYLYNLRAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLINDD 245
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
182 dkidifvhlleaakvlyqeyksnvevkikelnylktiqlkldfkknnfvgiadlst 241

Qy 246 TKKDMGLKLLSTGLVOIFPN--TIISLKIEKFDMLNIHQHCVKQCPENSGCFRHL 302
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
242 nhnnliltflsgmv--fenlaktvlsnlgndlgmqlqisghqcvkqcpngscfchl 299

Qy 303 DERECCKLLNYKQEDKCEENPNPTCNENNGCGDADATCTEEDSGSRKKITCCTCPD 362
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
300 dereeckllnykqgdgcvenpntcnenngcgdadakcteedsngkkitcectkpd 359

Qy 363 SYPLFDGIFCS 373
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
360 syplfdgifcs 370

RESULT 7
AA05834
ID AA05834 standard; Protein; 376 AA.
XX
AC AA05834;
XX
DT 02-AUG-1999 (first entry)
XX
DE Modified merozoite surface protein MSP-1-42.
```

```
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; mutant.
XX
OS Plasmodium falciparum.
OS Synthetic.
XX
PH Key Location/Qualifiers
FT Peptide 1..15
FT /note= "beta-casein signal peptide"
FT Peptide 371..376
FT /note= "6xHis tag"
FT Misc-difference 197
FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
FT Misc-difference 278
FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX
PN WO9920766-A2.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22225.
XX
PR 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX
PA (GENZ ) GENZYME TRANSGENICS CORP.
XX
PI Chen LH, Meade H;
XX
DR WPI; 1999-302742/25.
DR N-PSDB; AAX25593.
XX
PT New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
XX
PS Example; Fig 11; 43pp; English.
XX
CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The sequence has been modified to include
CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC tag. In addition, N181Q and N262Q mutations have been introduced to
CC eliminate 2 N-glycosylation sites. These modifications allow the
CC MSP-1-42 protein to be expressed in the milk of transgenic mice.
CC The invention generally relates to modified recombinant nucleic
CC acid sequences and methods for increasing the mRNA levels and
CC protein expression of proteins that are difficult to express in
CC cell culture systems, mammalian cell culture systems or in
CC transgenic animals. The preferred difficult protein candidates for
CC expression are those derived from lower organisms such as parasites,
CC bacteria and viruses that have DNA coding sequences of high AT
CC content or which have mRNA instability motifs or rare codons
CC relative to the recombinant expression system to be used.
XX
SQ Sequence 376 AA;

Query Match 56.0%; Score 1149.5; DB 20; Length 376;
Best Local Similarity 58.8%; Pred. No. 4e-73;
Matches 218; Conservative 53; Mismatches 75; Indels 25; Gaps 3;

Qy 6 MDNILSGFENEYDVYLYKPLAGVYRSKQIEKNIITFNLMNDILNSRLKRRYFLDLV 65
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
22 idnllskieneyevilykplagvyrslkqlennvmtfnvkvdklnsrfrnkrenfknl 81

Qy 66 ESDLMOQKHISSENYEYIIEDESKLLNSEQKNILLKSKYKIKESVENDIKFAQEGISYEKV 125
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
82 esdlipykdltsnyvkvdkpykflnkekrkflsynyikdsidtdinfandvgygkil 141

Qy 126 LAKYKDDLESIKKVIKEKEKFPSPPTTPPSPAKTDBQKESKFLPLTNIETLYNNLV 185
```



```
XX SQ Sequence 262 AA;
Query Match 31.38; Score 641.5; DB 22; Length 262;
Best Local Similarity 47.18; Pred. No. 1.3e-37;
Matches 131; Conservative 51; Mismatches 71; Indels 25; Gaps 3;
Qy 6 MDNLSGFENEYDVYLYKPLAGVYRSKQJTEKNITFNLNLDILNSRLKRRKRYELDVL 65
Db 7 IDNLSKIENEYVLYKPLAGVYRSKQJTEKNITFNLNLDILNSRLKRRKRYELDVL 66
Qy 66 ESDLQMFKHSSNEYIIDSFKLLNSEOKNILLKSYKIKESVENDIKFAQEGISYVEKV 125
Db 67 ESDLIPYKDLSSNVVVDKPYKFLNKEKDKFLSSYKIDSDIDTDFNFDVGLYKIL 126
Qy 126 LAKYKDDLESIKKVIKEKEKFPSPPTPPSPAKTDQKKESKFLPLTWIETLYNNLV 185
Db 127 sekysdlsdkyi-----ndkqgenekylypfinnietlyktvn 166
Qy 186 NKIDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKITDLFKNTNDFEAIKKLINDD 245
Db 167 dkidifvhlhlaekylytyeksnvevkielnylktiqlkdiadfkknunfvgiadlstdy 226
Qy 246 TKDMLGKLLSTGLVQIFPN--TIISKLEKGFQDML 280
Db 227 nhnlltkflstgmv--fenlaktvlsnldgnlqgm 262

RESULT 10
AAW36103
ID AAW36103 standard; Protein: 116 AA.
XX AC AAW36103;
XX DT 25-MAR-1998 (first entry)
XX DE PfMSPI(p19)A protein sequence.
XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX OS Plasmodium falciparum.
XX FH Key Location/Qualifiers
XX FT Region 1..95 /note= "amino acids derived from P. falciparum MSP1 p19
FT FT fragment"
FT FT Region 96..116 /note= "glycosylphosphatidylinositol anchoring sequence"
XX PN WO9730158-A2.
XX PD 21-AUG-1997.
XX PF 14-FEB-1997; 97WO-FR00290.
XX PR 14-FEB-1996; 96FR-0001822.
XX PA (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX DR WPI; 1997-425033/39.
XX DR N-PSDB; AAT94550.
XX XX Recombinant protein containing the merozoite surface protein-1 p19
PT fragment - useful in anti-malarial vaccines, diagnosis and protein
PT purification
XX PS Disclosure; Fig 1B; 85pp; French.
```

```
PS Disclosure; Fig 1B; 85pp; French.
XX CC This is the amino acid sequence of a recombinant protein comprising
CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX SQ Sequence 116 AA;
Query Match 30.08; Score 615; DB 18; Length 116;
Best Local Similarity 96.58; Pred. No. 3.2e-36;
Matches 110; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 281 NISQHCQVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCEENPNTCENNGGCCDADA 340
Db 3 NISQHCQVKKQCPENSGCFRHLDERECKCLLNYKQEGDKCEENPNTCENNGGCCDADA 62
Qy 341 TCTEEDSGSSSKKKTCTECTKPDSPYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 63 kcteedsngskkitcctckpdsyplfdgifcsssnflgisflilmlilysfi 116

RESULT 11
AAW22592
ID AAW22592 standard; Protein: 116 AA.
XX AC AAW22592;
XX DT 25-MAR-1998 (first entry)
XX DE PfMSPI(p19)A protein sequence.
XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX OS Plasmodium falciparum.
XX FH Key Location/Qualifiers
XX FT Region 1..95 /note= "amino acids derived from P. falciparum MSP1 p19
FT FT fragment"
FT FT Region 96..116 /note= "glycosylphosphatidylinositol anchoring sequence"
XX PN WO9730159-A2.
XX PD 21-AUG-1997.
XX PF 14-FEB-1997; 97WO-FR00291.
XX PR 14-FEB-1996; 96FR-0001821.
XX PA (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX DR WPI; 1997-425034/39.
XX DR P-PSDB; AAW22592.
XX XX Recombinant protein containing Plasmodium merozoite surface
PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
PT antibodies for diagnosis and protein purification
XX PS Disclosure; Fig 1B; 85pp; French.
```

CC This is the amino acid sequence of a recombinant protein comprising
CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
CC protein 1 (MSP-1) 19 kD C-terminal fragment (p19), linked to a
CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
CC The recombinant protein can be used for the production of anti-malarial
CC vaccines, where the p19 fragment provides a high level of protective
CC immunity since it includes epitopes not presented in the p42 fragment.
XX
SQ Sequence 116 AA;

Query Match 30.0%; Score 615; DB 18; Length 116;
Best Local Similarity 96.5%; Pred. No. 3.2e-36;
Matches 110; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 281 NISQHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCEENPNTCNNGGCDADA 340
DB 3 nisqhcvcvkqcpnsgcfrhldececkcllnkqegdkcvenpnptcnenngcdada 62
QY 341 TCTEEDSGSSRKKITCCTKPDSPYPLFDGIFCSSLNFGISFLILMLILYSFI 394
DB 63 kcteedsngskktctcktpdpsylfdgfcssnflgisflilmlilysfi 116

RESULT 12
AAB37609
ID AAB37609 standard; Protein; 108 AA.
XX
AC AAB37609;
XX
DT 27-FEB-2001 (first entry)
XX
XX Merozoite surface protein-119.
DE
XX Merozoite surface protein; protozoacide; vaccine; malaria.
KW
XX Plasmodium falciparum.
OS
XX WO200063245-A2.
PN
XX 26-OCT-2000.
PD
XX 20-APR-2000; 2000WO-GB01558.
PF
XX 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
PI WPI; 2001-015762/02.
XX
DR Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria -
XX
XX Example 5; Fig 15; 126pp; English.
PS
XX The present invention relates to non-natural variants of a C-terminal
XX fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is MSP-119 protein.
XX
SQ Sequence 108 AA;

Query Match 25.8%; Score 529; DB 22; Length 108;
Best Local Similarity 88.8%; Pred. No. 3.3e-30;
Matches 93; Conservative 4; Mismatches 4; Indels 4; Gaps 1;

QY 272 IEKFDQMLNISOHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCEENPNTCNE 331
DB 8 iegrh----niqhcvcvkqcpnsgcfrhldececkcllnkqegdkcvenpnptcne 63
QY 332 NNGGCDADATCTEEDSGSSRKKITCCTKPDSPYPLFDGIFCSSN 376
DB 64 nnggcdadakcteedsngskktctcktpdpsylfdgfcssn 108

RESULT 13
AAB37608
ID AAB37608 standard; protein; 96 AA.
XX
AC AAB37608;
XX
DT 27-FEB-2001 (first entry)
XX
XX Merozoite surface protein-1.
DE
XX Merozoite surface protein; protozoacide; vaccine; malaria.
KW
XX Plasmodium falciparum.
OS
XX WO200063245-A2.
PN
XX 26-OCT-2000.
PD
XX 20-APR-2000; 2000WO-GB01558.
PF
XX 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
PI WPI; 2001-015762/02.
XX
DR Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria -
XX
XX Example 2; Page 48; 126pp; English.
PS
XX The present invention relates to non-natural variants of a C-terminal
XX fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The present sequence is the
CC wild-type MSP-1 protein. This sequence was used to generate the variants
CC of the present invention. The non-natural variants of the present
CC invention are useful for immunising a mammal against malaria, and can be
CC used to treat malaria.
XX
SQ Sequence 96 AA;

Query Match 25.7%; Score 528; DB 22; Length 96;
Best Local Similarity 94.8%; Pred. No. 3.3e-30;
Matches 91; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 281 NISQHCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCEENPNTCNNGGCDADA 340
DB 1 nisqhcvcvkqcpnsgcfrhldececkcllnkqegdkcvenpnptcnenngcdada 60

QY 341 TCTEEDSGSSRRKKTCTCKPDSYPLFDGIFCSSN 376
 D 61 kcteedsngskktctckpdsyplfdgifs 96

RESULT 14
 AAW22593
 ID AAW22593 standard; Protein; 127 AA.
 XX AC AAW22593;
 XX 25-MAR-1998 (first entry)
 XX PFMSP1(p19)S protein sequence.
 KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
 XX Chimeric - Plasmodium vivax.
 OS Chimeric - Plasmodium falciparum.
 XX Key Location/Qualifiers
 FT Peptide 1..19 /note= "signal peptide"
 FT Protein 20..127 /note= "mature protein"
 FT Region 1..32 /note= "derived from P. vivax MSP1"
 FT Region 33..34 /note= "encoded by restriction enzyme sequence used to create the chimeric sequence"
 FT Region 35..127 /note= "derived from P. falciparum C-terminal p19 fragment of MSP1"
 PN WO9730159-A2.
 XX 21-AUG-1997.
 PD 14-FEB-1997; 97WO-FR00291.
 PF 14-FEB-1996; 96FR-0001821.
 XX (INSP) INST PASTEUR.
 PA (UUNY) UNIV NEW YORK STATE.
 XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX WPI; 1997-425034/39.
 DR P-PSDB; AAW22592.
 XX Recombinant protein containing Plasmodium merozoite surface protein-1 p42 fragment - useful in antimalarial vaccines, also new antibodies for diagnosis and protein purification
 PS Disclosure; Fig 1C; 85pp; French.
 XX This is the amino acid sequence of a chimeric protein comprising amino acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1) linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species. The recombinant protein can be used for the production of anti-malarial vaccines, where the p19 fragment provides a high level of protective immunity since it includes epitopes not presented in the p42 fragment.
 XX Sequence 127 AA;

QY 278 DMLNISOHCVKKOCPENSGCFRLHDERECKCLLNKQEGDKCEENPNTCNENNGSCD 337
 D 32 defnlsqvcvkqcpensgcfhldereckcllnkqegdkcvenpnptcnenngscd 91

QY 338 ADATCTEEDSGSSRRKKTCTCKPDSYPLFDGIFCS 373
 D 92 adakteedsngskktctckpdsyplfdgifs 127

RESULT 15
 AAW36102
 ID AAW36102 standard; Protein; 127 AA.
 XX AC AAW36102;
 XX 25-MAR-1998 (first entry)
 XX PFMSP1(p19)S protein sequence.
 KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
 XX Chimeric - Plasmodium vivax.
 OS Chimeric - Plasmodium falciparum.
 XX Key Location/Qualifiers
 FT Peptide 1..19 /note= "signal peptide"
 FT Protein 20..127 /note= "mature protein"
 FT Region 1..32 /note= "derived from P. vivax MSP1"
 FT Region 33..34 /note= "encoded by restriction enzyme sequence used to create the chimeric sequence"
 FT Region 35..127 /note= "derived from P. falciparum C-terminal p19 fragment of MSP1"
 PN WO9730158-A2.
 XX 21-AUG-1997.
 PD 14-FEB-1997; 97WO-FR00290.
 PF 14-FEB-1996; 96FR-0001822.
 XX (INSP) INST PASTEUR.
 PA (UUNY) UNIV NEW YORK STATE.
 XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX WPI; 1997-425033/39.
 DR N-PSDB; AAT94549.
 XX Recombinant protein containing the merozoite surface protein-1 p19 fragment - useful in anti-malarial vaccines, diagnosis and protein purification
 PS Disclosure; Fig 1C; 85pp; French.
 XX This is the amino acid sequence of a chimeric protein comprising amino acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1) linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
 CC p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species. The recombinant protein can be used for the production of anti-malarial vaccines, where the p19 fragment provides a high level of protective immunity since it includes epitopes not presented in the p42 fragment.
 XX Query Match 25.4%; Score 521; DB 18; Length 127;
 XX Best Local Similarity 93.8%; Pred. No. 1.5e-29;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 34.41 seconds
(without alignments)
235.763 Million cell updates/sec

Title: US-09-500-376-2

Perfect score: 2062

Sequence: 1 AISVTMDNLSGFENEDVI.....SNLIGISFLILMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	301	14.6	53	1 US-08-290-919-3	Sequence 3, Appli
2	282	13.7	53	1 US-08-290-919-4	Sequence 4, Appli
3	277	13.4	48	1 US-08-290-919-12	Sequence 12, Appl
4	274	13.3	48	1 US-08-290-919-2	Sequence 2, Appli
5	271	13.1	48	1 US-08-290-919-1	Sequence 1, Appli
6	266.5	12.9	106	1 US-08-290-919-11	Sequence 11, Appl
7	129.5	6.3	1494	3 US-08-755-587-186	Sequence 186, App
8	123.5	6.0	798	2 US-07-728-215-30	Sequence 30, Appl
9	123	5.8	740	1 US-08-257-073-5	Sequence 5, Appli
10	119	5.8	677	3 US-08-480-640A-115	Sequence 115, App
11	119	5.8	677	3 US-08-480-640A-193	Sequence 193, App
12	119	5.8	677	3 US-08-295-802-115	Sequence 115, App
13	119	5.8	677	4 US-08-686-968C-58	Sequence 58, Appl
14	119	5.8	677	4 US-08-686-968C-193	Sequence 193, App
15	119	5.8	677	4 US-08-488-237A-115	Sequence 115, App
16	119	5.8	677	4 US-08-488-237A-193	Sequence 193, App
17	118	5.7	1786	4 US-08-973-462-8	Sequence 8, Appli
18	117.5	5.7	414	5 PCT-US93-03077-3	Sequence 3, Appli
19	117.5	5.7	984	1 US-08-257-073-3	Sequence 3, Appli
20	117.5	5.7	984	2 US-08-184-009-120	Sequence 120, App
21	117.5	5.7	984	2 US-08-458-356-120	Sequence 120, App
22	117.5	5.7	1093	5 PCT-US93-03077-1	Sequence 1, Appli
23	116	5.6	540	1 US-08-286-325A-2	Sequence 2, Appli
24	116	5.6	3135	1 US-08-323-170B-2	Sequence 2, Appli
25	115.5	5.6	2710	1 US-08-480-604A-6	Sequence 6, Appli
26	115.5	5.6	2710	2 US-08-405-496A-6	Sequence 6, Appli
27	115	5.6	984	1 US-08-242-932-2	Sequence 2, Appli

Sequence 2, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 14, Appli
Sequence 14, Appli
Sequence 14, Appli
Sequence 14, Appli
Sequence 4, Appli
Sequence 184, App
Sequence 44, Appl
Sequence 52, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-08-290-919-3

; Sequence 3, Application US/08290919

; Patent No. 5720959

; GENERAL INFORMATION:

; APPLICANT: HOLDER, ANTHONY A.

; APPLICANT: BLACKMAN, MICHAEL J.

; APPLICANT: CHAPPEL, JONATHAN A.

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA

; TITLE OF INVENTION: VACCINE

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DABRY & CUSHMAN, L.L.P.

; STREET: 1100 NEW YORK AVENUE, N.W.

; CITY: WASHINGTON

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3918

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/290,919

; FILING DATE: 04-OCT-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9203821.5

; FILING DATE: 22-FEB-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB93/00367

; FILING DATE: 22-FEB-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: KOKULIS, PAUL N.

; REGISTRATION NUMBER: 16,773

; REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 861-3000

; TELEFAX: (202) 822-0944

; TELEX: 6714627 CUSH

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 53 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FEATURE:

; NAME/KEY: Modified-site

[illegible]

```

CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/JWL/6BCB/
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= X
OTHER INFORMATION: /note= "X = M and N, Or N"
US-08-290-919-1

Query Match 13.1%; Score 271; DB 1; Length 48;
Best Local Similarity 97.9%; Pred. NO. 7.2e-14;
Matches 46; Conservative 1; Mismatches 0; Indels

Qy 282 ISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 328
2 ISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 48

Db 2 ISOHCVKKQCPENSGCFRHLDERECKLLNYKQEGDKCVENPNT 48

RESULT 6
US-08-290-919-11
Sequence 11, Application US/08290919
Patent No. 5720959
GENERAL INFORMATION:
APPLICANT: HOLDER, ANTHONY A.
APPLICANT: BLACKMAN, MICHAEL J.
APPLICANT: CHAPPEL, JONATHAN A.
TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO MALARIA
TITLE OF INVENTION: VACCINE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,919
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9203821.5
FILING DATE: 22-FEB-1992
PRIOR APPLICATION DATA: PCT/GB93/00367
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 212242/HCM/MJL/6BC8/
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 106 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-290-919-11

Query Match 12.9%; Score 266.5; DB 1; Length 106;
Best Local Similarity 49.0%; Pred. No. 4.2e-13;
Matches 50; Conservative 17; Mismatches 32; Indels 3; Gaps 3;
QY 278 DMLNIS-OHOCV-KKQCPENSGCFRHLDERECKCLLNYKQ-EGDKCVENPNPCNENNG 334
Db 5 DLLGVDPKHVCVTRDIPKNAGCFRDNGTEFWFCLLGYKKGGTCVNNNPTCDING 64
QY 335 GCDADAKTEDSGNGKKTCTCECTKPDSPFLDFGIFCSSN 376
Db 65 GCDPTASQNAESTENSKKIICTKCTPPNAYYEGVFCSSS 106

RESULT 7
US-08-755-587-186
Sequence 186, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
TITLE OF INVENTION: susceptibility gene and uses thereof.
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 186:
SEQUENCE CHARACTERISTICS:
LENGTH: 1494 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-755-587-186

Query Match 6.3%; Score 129.5; DB 3; Length 1494;
Best Local Similarity 22.9%; Pred. No. 0.17;
Matches 95; Conservative 68; Mismatches 167; Indels 85; Gaps 22;
QY 4 VTMDNILSGFEN-----EYDVLYLKPLAGVYRSLKKQIEKNIFTFFNLNLN---DILNSR 54
Db 760 VTMSNKQOLTANTGNTKDFDFYL-----SFQASRKNIRVRESLNKARSLNQK 811
QY 55 LKKRKY--FLDVLESIDLMOFKHIS----SNEYIITEDSKLLNSQKNTLLKSYKYIKESV 108
Db 812 WTEELNFSLSNSELLPGIDIKKTDISNHEVLE-----NTERKDKITKESLIGTENI 865
QY 109 -----ENDIKFAQEG--ISYVE---KVLAKYKDDLESIKKVIKEKEKFPSPPTTP 155
Db 866 LLILQRPESKIRKIKESAVLGFHTASGKKIETITKESLDKVKNLF-EKEQDNSEITNFS 924
QY 156 PPAKTDQEKESK-----FLPFLNIETLYNNL-VNKIDDYLINLAKAKINDCNV 204
Db 925 HRGAKMSKRECKDGRRELACGTEITTTPEYETHSLEKKKVSNEIAALRPRLSDNL 984
QY 205 EKDEAHVKITKLSDLKAIDDKIDLFKNHNDFAIKKLINDDTKKDKMLGKLLSLGLVQNF 264
Db 985 YKQENLKISDHA-----SQKVDVHEN-TEKETAK-----KPTMYTNQSTYSAIENSP 1031
QY 265 NTIISKLEIGKFDMLNISQHCVKKOCPENSGCFRHLDERECKCLLNYKQEGDKCVEN 324
Db 1032 LT-FTQDTEERF-----SVSEASLFEAKKWLREGEDDDQSERINA-AKVNCKEYPDDYVE- 1084
QY 325 PNPTCNENGGCDADAKTEED---SGSNGKKTCTCECTKPDSPFLDFGIFCSSS 375
Db 1085 -NPSC-----GNSNSAITENDKNHLSKOGSTYLSNSTMNSYSYHPG-FCHSS 1132

RESULT 8
US-07-728-215-30
Sequence 30, Application US/07728215
Patent No. 5962643
GENERAL INFORMATION:
APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25


```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/728,215
; FILING DATE: 19910711
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P31 8717
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-728-215-30

Query Match 6.0%; Score 123.5; DB 2; Length 798;
Best Local Similarity 20.8%; Pred. No. 0.21;
Matches 103; Conservative 67; Mismatches 146; Indels 179; Gaps 29;

QY 48 NDILNSR----LKKRYFLDVLSDLMQFHHISSNEVIYIEDSFKLLNSEQKNTLLSKYK 103
Db 78 DDLENPRGSKDKKKNVNRSGTAELKPEDIHQIOPQOLVLRSGEPQFTTLKFKR 137
QY 104 IKESVENDIKFAQEGISYIEKVLAKYKDDLESIKV---IKEKEKPPSS----- 150
Db 138 -AEDYPIDLYLMD-LSY-----SMKDDLENYKSLGTDLMNEMRRITSDFRIGFGSFE 189
QY 151 ---PPTTPPSPAK-----TDEQKESKF---LPFTNIETLYNNLVNK-----ID--- 189
Db 190 KTWMPYISTTPAKLRNPCTSEQNTTFFSYKKNVLSLNKGEVFNELVKGORISGNLDSPE 249
QY 190 ---DYLNL-----KAKI-----ND---CNVEKDE-- 208
Db 250 GGFDAIMQAVCGSLIGWRNVTRLLVFSTDAFHFAGDKLGGIVLPNDGQCHLENNMT 309
QY 209 -----AHVKITKLSLKDIDKIDLFKNHNDPEAKKLINDTKDMLGKLLSTG 258
Db 310 MSHYDYPSTAHL-VOKLSE-----NNITQIFAVTEEPQYVKELNLPKSAVGTLSANS 364
QY 259 -----LVQNEPNTIISKLI--EGKFODMLNISQHCVK-----289
Db 365 SNVIQLIIDAYNSLSSEVILENKLSEGVTSYKCKNGVNGTGENGRKCSNISIGDEV 424
QY 290 -----KQCP-ENSGCFR-----HLDERE-----ECKCLLNYKQEGDKCVENPNPT 328
Db 425 QFEISITSNCRPKKSDSFIRPLGFTGEEVILQYICECEQSEGIPSPKCHEG-NGT 483
QY 329 -----CNENGG---C-----DADAKTEEDS-----GSNGKKI--TCECIPD- 362
Db 484 FEGACACRNGRVRGRHCECSTDEVNSDMDAYCRKENSEISCSNNGBCVCGQCVCRKRN 543
QY 363 SYPLFDGIFCSSNF 377
Db 544 TNEYSGKFCECDNF 558

RESULT 9
US-08-257-073-5
; Sequence 5, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P. C.
```

```

; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 740 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
;
US-08-257-073-5

Query Match 6.0%; Score 123; DB 1; Length 740;
Best Local Similarity 22.9%; Pred. No. 0.21;
Matches 96; Conservative 53; Mismatches 110; Indels 160; Gaps 22;

QY 47 LNDILNSRLKRR---KYFLDVLSDLMQFHHISSNEVI-----IEDSFKLLNSEQ----- 93
Db 289 VNDVLAVNMKNMKTGNFMDFDVKSTLAQSGGLGSLNDLINFNLQGEIGENLLNITKMLNG 348
QY 94 -KNTL-----LKSXYI---KESVENDIK-----FAQEGISY-----YE 123
Db 349 DKNLESFPLDQLNMLKDNLINEFILLNKTSVLNKLDKLLLLLYKAYSVKKRAQE 408
QY 124 KVL-----AKYKDDLE-----SIKKVIKEKEK-FPSSPPTTPPSPAKT 162
Db 409 KGLPEPTVTNEEYVEELKGLDMGKILKFSKSLKLLKLNKIFPKKEDNQAVDTKSM 468
QY 163 EOKKESKFLPFL-----TNIETLYNNLVNKIDDDYLLINKAKINDCNV----- 204
Db 469 EEPK-KQAOPALRGVEPTSDSNMNSINNYMDEIDFKELEIENNNTPNVPPVTSQSKKN 527
QY 205 -----EKDEAHVKITKLS-DLKAIDDKIDLFKNHNDFE 236
Db 528 KNETVSGMDENFDNHPENYKFEYYDENDMDMEVKVKIGVTLK----KPELKNGVSE 583
QY 237 AIK-----KLINDTK--KDMGLKLLSTGLVQNPNTIISKLEGRFQ----- 277
Db 584 TIKLIHLGNKDKKHIEAINDIOIKQLAIYNE--LMVYNG--NKNIQOIFQOONILE 639
QY 278 -DMLNISQHCQVVKQ-----CPENSGCFRHLDEREECKCLLNYKQEGDK 320
Db 640 NDVLNQETEEEMEKQVEAITKQIEAEVDALAPKNK-----EEEEKEKEKEKEKEK 692
```

RESULT 10
US-08-480-640A-115
; Sequence 115, Application US/08480640A
; Patent No. 6033904
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,640A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; IMMEDIATE SOURCE:
; CLONE: 515-85.1
; POSITION IN GENOME:
; MAP POSITION: -23.2
; UNITS: %G
US-08-480-640A-115

Query Match 5.8%; Score 119; DB 3; Length 677;
Best Local Similarity 22.2%; Pred. No. 0.37;
Matches 76; Conservative 65; Mismatches 137; Indels 64; Gaps 18;
QY 18 DVYILKPLAGVYSLKKQIEKNITFF-----NLNLDILNLSRLKKRYFLDVL-----SDLM 70
DB 185 DAFHIPHTHLSRL-RDINRIELLLKKYPNNIIDIYSDSIKSNSSFIHLMIITSNMF 243
QY 71 QFKHISSEYI--IEDSFKLLNSEQ-KNTLLKSYKY-----IKESVENDIKFAOEGISYE 123
DB 244 PAIIPSVNDIFSTVVDRLINMYGKCVAMFSDINMIDLESDDSDYIFIEKNISIYD 303
QY 124 KVLAKYKDDLESIKVYKEKEKFPSPPTPPSPARTDQKKESKFLPLTNIETLYNN 183
DB 304 ---VKCRDFANMIRDKVKREKNRILTT-----KCEDIIRYIKLFS-----KNRIND 346
QY 184 LVNKIDYLLNKAINDCNVKEDEAHVITKLSDLKAIDDKIDLKNNHNDFAIKKLIN 243
DB 347 ENNKVEEVLTHID-----NVSKNNK-----LSLSDISSLMDQFRL-----NPCTIRNILLS 392
QY 244 DDTKDKMLGKLLSTGLVQNF-----PNTIISKLEIGKF-ODMLNISOHCQVK--KQCPEN 295
DB 393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKIKGVIVMDVDYISTNLKLYHKQLYDK 449

QY 244 DDTKDKMLGKLLSTGLVQNF-----PNTIISKLEIGKF-ODMLNISOHCQVK--KQCPEN 295
DB 393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKIKGVIVMDVDYISTNLKLYHKQLYDK 449
QY 296 SGCFRHLDERECKC-----LLNYKQEGDKCVENPNPTCNE 331
DB 450 MSTFEYKRDICKCKSCISDSITHHIVETTSCI-NYKSTDN 490
RESULT 11
US-08-480-640A-193
; Sequence 193, Application US/08480640A
; Patent No. 6033904
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,640A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-480-640A-193

Query Match 5.8%; Score 119; DB 3; Length 677;
Best Local Similarity 22.2%; Pred. No. 0.37;
Matches 76; Conservative 65; Mismatches 137; Indels 64; Gaps 18;

QY 18 DVYILKPLAGVYSLKKQIEKNITFF-----NLNLDILNLSRLKKRYFLDVL-----SDLM 70
DB 185 DAFHIPHTHLSRL-RDINRIELLLKKYPNNIIDIYSDSIKSNSSFIHLMIITSNMF 243
QY 71 QFKHISSEYI--IEDSFKLLNSEQ-KNTLLKSYKY-----IKESVENDIKFAOEGISYE 123
DB 244 PAIIPSVNDIFSTVVDRLINMYGKCVAMFSDINMIDLESDDSDYIFIEKNISIYD 303
QY 124 KVLAKYKDDLESIKVYKEKEKFPSPPTPPSPARTDQKKESKFLPLTNIETLYNN 183
DB 304 ---VKCRDFANMIRDKVKREKNRILTT-----KCEDIIRYIKLFS-----KNRIND 346
QY 184 LVNKIDYLLNKAINDCNVKEDEAHVITKLSDLKAIDDKIDLKNNHNDFAIKKLIN 243
DB 347 ENNKVEEVLTHID-----NVSKNNK-----LSLSDISSLMDQFRL-----NPCTIRNILLS 392
QY 244 DDTKDKMLGKLLSTGLVQNF-----PNTIISKLEIGKF-ODMLNISOHCQVK--KQCPEN 295
DB 393 SATIK---SKLLALRAVKNWKCYSLTNVSMYKIKGVIVMDVDYISTNLKLYHKQLYDK 449

; CURRENT FILING DATE: 1996-07-25
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 193
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Swinepox virus
US-08-686-968C-193

Query Match 5.8%; Score 119; DB 4; Length 677;
Best Local Similarity 22.2%; Pred. No. 0.37;
Matches 76; Conservative 65; Mismatches 137; Indels 64; Gaps 18;

QY 18 DVIYKPLAGVYSLKKQIEKNITFF-----NLNLDILNSRLKRRKYFLDVLVLE---SDLM 70
DB 185 DAFHIPPTHTSLRSL-RDINRIIELLKKYPNNNIIDVSDSIKSNSSFIHILHMIISNMF 243
QY 71 QFKHISSEYI--IEDSFKLINSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
DB 244 PAIIPSVNDFISTVVDKRLINMYGKCVAMFSYDINMIDLESDDSDYIFIEKNISIYD 303
QY 124 KVLAKYKDDLESIKKVIKEKEKPPSPPTPPSPAKTDQKKESKFLPFLTNIETLYNN 183
DB 304 ---VKCRDFANMIRDKVKREKNRILTT-----KCEDIIIRYIKLFS-----KNRIND 346
QY 184 LVNKIDDYLINLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNHNDFAIKKLIN 243
DB 347 ENNKVEEVLIHID-----NVSKNNK-----LSLSDISSLMDQFRL-----NPCTIRNILLS 392
QY 244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLEGRF-ODMLNISOHQCVK--KQCPEN 295
DB 393 SATIK---SKLLALRAVKNKCYSLTNVSMYKKGIVVMDVDYISTNLIKVKHQLYDK 449
QY 296 SGCFRHLDERECKC-----LLNKOEGDKCVENPNPTCNE 331
DB 450 MSTPEYKRDIKSKCKSCSDSIITHIYETTSCI-NYKSTND 490

RESULT 15

US-08-488-237A-115
; Sequence 115, Application US/08488237A
; Patent No. 6251403
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: Recombinant Swinepox Virus
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,237A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 115:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 677 amino acids

; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Swinepox virus
; STRAIN: Kasza
; INDIVIDUAL ISOLATE: S-SPV-001
; IMMEDIATE SOURCE:
; CLONE: 515-85.1
; POSITION IN GENOME:
; MAP POSITION: -23.2
; UNITS: 8C
US-08-488-237A-115

Query Match 5.8%; Score 119; DB 4; Length 677;
Best Local Similarity 22.2%; Pred. No. 0.37;
Matches 76; Conservative 65; Mismatches 137; Indels 64; Gaps 18;

QY 18 DVIYKPLAGVYSLKKQIEKNITFF-----NLNLDILNSRLKRRKYFLDVLVLE---SDLM 70
DB 185 DAFHIPPTHTSLRSL-RDINRIIELLKKYPNNNIIDVSDSIKSNSSFIHILHMIISNMF 243
QY 71 QFKHISSEYI--IEDSFKLINSEQ-KNTLLKSYKY----IKESVENDIKFAQEGISYYE 123
DB 244 PAIIPSVNDFISTVVDKRLINMYGKCVAMFSYDINMIDLESDDSDYIFIEKNISIYD 303
QY 124 KVLAKYKDDLESIKKVIKEKEKPPSPPTPPSPAKTDQKKESKFLPFLTNIETLYNN 183
DB 304 ---VKCRDFANMIRDKVKREKNRILTT-----KCEDIIIRYIKLFS-----KNRIND 346
QY 184 LVNKIDDYLINLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNHNDFAIKKLIN 243
DB 347 ENNKVEEVLIHID-----NVSKNNK-----LSLSDISSLMDQFRL-----NPCTIRNILLS 392
QY 244 DDTKKDMLGKLLSTGLVQNF-----PNTIISKLEGRF-ODMLNISOHQCVK--KQCPEN 295
DB 393 SATIK---SKLLALRAVKNKCYSLTNVSMYKKGIVVMDVDYISTNLIKVKHQLYDK 449
QY 296 SGCFRHLDERECKC-----LLNKOEGDKCVENPNPTCNE 331
DB 450 MSTPEYKRDIKSKCKSCSDSIITHIYETTSCI-NYKSTND 490

Search completed: August 8, 2001, 12:29:56
Job time: 44 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 60.73 Seconds
(without alignments)
393.312 Million cell updates/sec

Title: US-09-500-376-2

Perfect score: 2062

Sequence: 1 AISVTMDNLSGFENEYDVI.....SNFLGSLFLILMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 segs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred: No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1283.5	62.2	1639	19 AAW54145	P. falciparum synt
2	1191.5	57.8	355	20 AAY09372	Merozoite surface
3	1191.5	57.8	355	20 AAY05832	Merozoite surface
4	1191.5	57.8	361	20 AAY09373	Merozoite surface
5	1191.5	57.8	361	20 AAY05833	Merozoite surface
6	1184.5	57.4	376	20 AAY09374	Modified merozoite
7	1184.5	57.4	376	20 AAY05834	Modified merozoite
8	1124	54.5	1654	6 AAY05077	Sequence of the P1
9	651.5	31.6	262	22 AAB37610	Merozoite surface
10	640	31.0	116	18 AAW36103	PfMSP1(p19)A prote
11	640	31.0	116	18 AAW22592	PfMSP1(p19)A prote

12	554	26.9	108	22	AA37609	Merozoite surface
13	553	26.8	96	22	AA37608	Merozoite surface
14	546	26.5	127	18	AAW22593	PfMSP1(p19)S prote
15	546	26.5	127	18	AAW36102	PfMSP1(p19)S prote
16	307	14.9	54	14	AA41356	MSP1EGF2A EGF2-11k
17	288	14.0	54	14	AA41357	MSP1EGF2B EGF2-11k
18	282	13.7	49	14	AA41355	MSP1EGF1B EGF1-11k
19	279	13.5	49	14	AA41354	MSP1EGF1A EGF1-11k
20	279	13.5	93	22	AA37611	Merozoite surface
21	266.5	12.9	106	14	AA41358	P. yoelii combined
22	200.5	9.7	330	21	AA70278	Recombinant vaccin
23	180.5	8.8	980	21	AA18294	Plasmodium falcipa
24	138.5	6.7	1979	21	AA18171	Plasmodium falcipa
25	136	6.6	2573	21	AA18234	Plasmodium falcipa
26	133	6.5	1712	21	AA18205	Plasmodium falcipa
27	130	6.3	2500	21	AA18272	Plasmodium falcipa
28	129.5	6.3	1817	21	AA18255	Plasmodium falcipa
29	128	6.2	1254	11	AA07503	Merozoite apical-en
30	128	6.2	1254	18	AAW24575	Merozoite apical-e
31	127.5	6.2	1997	21	AA18287	Plasmodium falcipa
32	127	6.2	26	22	AA371974	Plasmodium falcipa
33	127	6.2	1166	21	AA18268	Plasmodium falcipa
34	125.5	6.1	2485	21	AA18172	Plasmodium falcipa
35	125	6.1	696	21	AA18181	Plasmodium falcipa
36	125	6.1	2206	21	AA18254	Plasmodium falcipa
37	125	6.1	2285	20	AAW98149	Bacillus subtilis
38	124	6.0	558	21	AA18273	Plasmodium falcipa
39	123.5	6.0	798	22	AA36937	Human integrin bet
40	123.5	6.0	963	19	AAW70540	Integrin beta-1 ch
41	123.5	6.0	1188	21	AA18183	Plasmodium falcipa
42	123	6.0	400	18	AAW55691	H. pylori ORF 06gp
43	123	6.0	407	21	AA18248	Plasmodium falcipa
44	123	6.0	740	13	AA27530	Plasmodium falcipa
45	123	6.0	740	16	AA668838	Plasmodium falcipa

ALIGNMENTS

RESULT 1	
AAW54145	
ID	AAW54145 standard; Protein; 1639 AA.
XX	AAW54145;
XX	AC
XX	23-SEP-1998 (first entry)
XX	DE P. falciparum synthetic gp190 protein.
XX	gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX	monoclonal antibody; passive immunisation; parasite.
XX	Plasmodium falciparum.
XX	Synthetic.
XX	WO9814583-A2.
XX	09-APR-1998.
XX	02-OCT-1997; 97WO-EP05441.
XX	02-OCT-1996; 96DE-4040817.
XX	(BUJA/) BUJAR H.
XX	Bujard H, Pan W, Tolle R;
XX	WPI; 1998-240088/21.
XX	N-PSDB; AAV21451, AAV35363.
XX	Recombinant production of complete gp190/MSP-1 Plasmodium surface
XX	protein - useful in anti-malaria vaccines, also stabilising genes by
XX	reducing their AT content

Example 1; Fig 3c; 48pp; German.

This sequence represents a modified plasmodium falciparum gp190/MSP-1 (merozoite surface) protein. The gene encoding this protein has been stabilised by reducing the AT content of the nucleotide sequence. Such a protein is useful in vaccines against malaria or for producing monoclonal antibodies (for passive immunisation). The complete gp190 protein can now be produced outside the parasite and has, at least over extended regions, the native pattern of folding. Larger amounts of the protein can be produced recombinantly than would be possible using the parasites as source.

XX Sequence 1639 AA;

SQ

Query Match 62.2%; Score 1283.5; DB 19; Length 1639;
Best Local Similarity 62.1%; Pred. No. 5.7e-82;
Matches 242; Conservative 73; Mismatches 75; Indels 21; Gaps 2;

QY 6 MDNITLGSFENEDYVILKPLAGVYRSLSKKQLEKNIFTEINLNLDNLNSRLKRRKYFLDVL 65
Db :||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1270 idnllskieneyevilykplagvyrslkkqlennvmtnfvnvkdlnsrnfknrenfknvl 1329

QY '66 ESDLMQRFKHISSEYIIDSFKFLNSEQNLTLLSKYIKESVENDIKFQAEGISYYEKV 125
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1330 esdlipykdltesnyvvkdpkyflnkekrdflesynykdsidtdinfandvgyykil 1389

QY 126 LAKYKDDLESHTKKVIKEKEFPSPPTPPSPAKTDQQKESKELPFLTNIETLYNNLV 185
Db ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1390 sekysdlsidsikyl-----ndkqgenekyflpfnnaletlyktvn 1429

QY 186 NKIDYILINKAKINDCNVEKDEAHVKITLSDLKAIIDKDLFNHNDFFAIIKLINDD 245
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1430 dkidlfiwhleakvlnytyeksnevkieinyktiqliadfknnmfvgiadlstdy 1489

QY 246 TKDKMLGKLLSTGLV-QNPFTYIISKIEGRFDMLNISQHCVKQKCPENGSGCFRHLDE 304
Db ::|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1490 nhnallkfistgmvfenaiktvtlslldgnlqgmlnisghcvkkgcpngsgcfrhde 1549

QY 305 REECKCLLNVKQEGDKCVENPNPTCNENNCGCDADAKCTEEDSGSNGKKITCECTKPDsy 364
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1550 reeckcllnvykqgdckvenpnptcnennnggcdadakcteedsgsngkkitcectkpdsy 1609

QY 365 PLFDGIFCSSNFILGISPLLIIMLILYSFI 394
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1610 plfdgifcssnfligifillimlilysfi 1639

RESULT 2
AAAY09372
ID AAAY09372 standard; Protein; 355 AA.
XX
AC AAAY09372;
AC
DT 31-AUG-1999 (first entry)
XX
XX Merozoite surface protein MSP-1-42.
XX
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
OS Plasmodium falciparum.
XX
XX WO9920774-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98WO-US22226.
XX
PF 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
PR

QY 66 ESDLMQPKHISSEYIIEDSFKLNSBQKNTLLKSYKIKESVENDIKFAQEGISYYEKV 125
 Db 67 esdlipkydltsnyvvykpykflnkekrdkflssynyikdsidtdinfandvlgvykil 126
 QY 126 LAKYKDLESITKKVKEKEKFPSPPTPPSPAKTDEQKESKFLPFLNIETLYNNLV 185
 Db 127 sekyksdlsikkyi-----ndkqgenekylpflnnietlyktvn 166
 QY 186 NKIDDLINLAKTDCNVKDEAHVKITKLSDLKATDDKIDLPKHNDFEATKKLINDD 245
 Db 167 dkidlfvihleakvinytyeksnvevkikelnlyktlqkldafkknntfvgadlstdy 226
 QY 246 TKDMLGKLLSTGLV-ONFPNTIISKLEKGFQDMLNISQHCVKKOCNPENSGCFRHLDE 304
 Db 227 nhnllckflstgmvfennlaktvlnldgnlqgmlnshqcvkqcpqnsqgcfrrhde 286
 QY 305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKTEEDSGSNKKITCECTKPDYS 364
 Db 287 reeckcllnykgqegdkcvenpnptcnennnggcdadakcteedsngngkkitcctkpdys 346
 QY 365 PLFDGIFCS 373
 Db 347 plfdgifcs 355
 RESULT 5
 AAY05833
 ID AAY05833 standard; Protein; 361 AA.
 XX
 AC AAY05833;
 XX
 DT 02-AUG-1999 (first entry)
 XX
 DE Merozoite surface protein MSP-1-42.
 XX
 KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9920766-A2.
 XX
 PD 29-APR-1999.
 XX
 PF 20-OCT-1998; 98WO-US22225.
 XX
 PR 15-MAY-1998; 98US-0085649.
 PR 20-OCT-1997; 97US-0062592.
 XX
 PA (GEN2) GENZYME TRANSGENICS CORP.
 XX
 PI Chen LH, Meade H;
 XX
 DR WPI; 1999-302742/25.
 DR N-PSDB; AAX25587.
 XX
 PT New modified recombinant nucleic acid sequences useful for producing
 PT malarial DNA vaccine
 XX
 PS Disclosure; Fig 2; 43pp; English.
 XX
 CC This present sequence comprises a 42 kDa C-terminal portion of
 CC malarial merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The C-terminal end of the sequence is
 CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
 CC AAX25593) encoding MSP-1-42 have been modified according to a method
 CC of the invention in order to improve expression in mammalian cells
 CC and in transgenic animals. The invention provides modified
 CC recombinant nucleic acid sequences and methods for increasing the
 CC mRNA levels and protein expression of proteins that are difficult
 CC to express in cell culture systems, especially mammalian cell

CC culture systems or in transgenic animals. The preferred difficult
 CC protein candidates for expression are those derived from lower
 CC organisms such as parasites, bacteria and viruses that have DNA
 CC coding sequences of high AT content or which have mRNA instability
 CC motifs or rare codons relative to the recombinant expression system
 CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.
 XX Sequence 361 AA;
 SQ
 Query Match 57.8%; Score 1191.5; DB 20; Length 361;
 Best Local Similarity 60.2%; Pred. No. 2.5e-76;
 Matches 222; Conservative 54; Mismatches 72; Indels 21; Gaps 2;
 QY 6 MDNLSGFENEYDVYILKPLAGVYRSLLKQIEKNIFFTNLDILNSRLKKRYFLDVL 65
 Db 7 idnlskieneyevlylkplagvyrsllkqlennvmtfnvkvdklnsrfrknfknvl 66
 QY 66 ESDLMQPKHISSEYIIEDSFKLNSBQKNTLLKSYKIKESVENDIKFAQEGISYYEKV 125
 Db 67 esdlipkydltsnyvvykpykflnkekrdkflssynyikdsidtdinfandvlgvykil 126
 QY 126 LAKYKDLESITKKVKEKEKFPSPPTPPSPAKTDEQKESKFLPFLNIETLYNNLV 185
 Db 127 sekyksdlsikkyi-----ndkqgenekylpflnnietlyktvn 166
 QY 186 NKIDDLINLAKTDCNVKDEAHVKITKLSDLKATDDKIDLPKHNDFEATKKLINDD 245
 Db 167 dkidlfvihleakvinytyeksnvevkikelnlyktlqkldafkknntfvgadlstdy 226
 QY 246 TKDMLGKLLSTGLV-ONFPNTIISKLEKGFQDMLNISQHCVKKOCNPENSGCFRHLDE 304
 Db 227 nhnllckflstgmvfennlaktvlnldgnlqgmlnshqcvkqcpqnsqgcfrrhde 286
 QY 305 REECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKTEEDSGSNKKITCECTKPDYS 364
 Db 287 reeckcllnykgqegdkcvenpnptcnennnggcdadakcteedsngngkkitcctkpdys 346
 QY 365 PLFDGIFCS 373
 Db 347 plfdgifcs 355
 RESULT 6
 AAY09374
 ID AAY09374 standard; Protein; 376 AA.
 XX
 AC AAY09374;
 XX
 DT 31-AUG-1999 (first entry)
 XX
 DE Modified merozoite surface protein MSP-1-42.
 XX
 KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.
 XX
 OS Plasmodium falciparum.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..15
 FT /note= "beta-casein signal peptide"
 FT Peptide 371..376
 FT /note= "6xHis tag"
 FT Misc-difference 197
 FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
 FT Misc-difference 278
 FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
 XX
 PN WO9920774-A2.

```

XX PD 29-APR-1999.
XX KW 20-OCT-1998; 98WO-US222226.
XX PF
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 97US-0062592.
XX PA (GENZ ) GENZYME TRANSGENICS CORP.
XX PI Chen LH, Meade H;
XX FT WPI; 1999-288313/24.
XX DR P-PSDB; AAX56008.
XX PR
XX PT Modified malarial protein for use in anti-malarial vaccines
XX PS Example; Fig 11; 35pp; English.
XX CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The sequence has been modified to include
CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC tag. In addition, N181Q and N262Q mutations have been introduced to
CC eliminate 2 N-glycosylation sites. These modifications allow the
CC MSP-1-42 protein to be expressed in the milk of transgenic mice.
XX CC
XX SQ Sequence 376 AA;

Query Match 57.4%; Score 1184.5; DB 20; Length 376;
Best Local Similarity 59.9%; Pred. No. 8.1e-76;
Matches 221; Conservative 53; Mismatches 74; Indels 21; Gaps 2;

QY 6 MDNLGFEYEDVYIKPLAGVYRSLLKQIEKNIFTNLDILNSRLKRRKRYFDVL 65
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
22 idnllskieneyevlykplagvyrslkqlennvmtfnvkvdkilnsrlnkrenfknvl 81

QY 66 ESDLMOFKHSSNEYIIEDSFKLNSQKNTLLKSKYIKESVENDIKFAQEGISYVEKV 125
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
82 esdlipkydltssnyvkvdkpkyfinkkrdkfissynpykdsidtdinfandvgyykil 141

QY 126 LAKYKDDLESIKKVIKEEKEKFPSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
142 sekyksdldsiiky|-----ndkqgenekylpflnnietlyktvn 181

QY 186 NKIDYDYLINLAKINDCNVEKDEAHVKITKLSDLKATIDDKIDLFKNHNDPEATKKLINDD 245
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
182 dkidlfvihleakvlytyeksnvevkikelnylktlqdkladfkknntfvgiadlstly 241

QY 246 TKKMDLGLKLLSTGLV-QNFPTTIISKLEGFQDMLNISQCVKKQCPENSGCFRHLDE 304
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
242 nhnllckflstgmvfennlaktvlnlildnlgmlqldshqcvkkqcpnsgcfhrhlde 301

QY 305 REECKCLLNKQEGDKCVENPNPTTCNENNGCGDADAKTEEDSGNGKKITCTCKTPDSY 364
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
302 reeckcllnlykqgdckvenpnptcnenngcgdadakcteedsnggkkitcctcktpdsy 361

QY 365 PLFDGIFCS 373
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
362 plfdgifcs 370

RESULT 7
AAV05834
ID AAY05834 standard; Protein; 376 AA.
XX AC AAY05834;
XX XX
XX DT 02-AUG-1999 (first entry)
XX DE Modified merozoite surface protein MSP-1-42.

```

```

XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal; mutant.
XX
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Peptide 1..15
XX FT /note= "beta-casein signal peptide"
XX FT Peptide 371..376
XX FT /note= "6xHis tag"
XX FT Misc-difference 197
XX FT /note= "Asn in native MSP-1-42 (N181Q mutation)"
XX FT Misc-difference 278
XX FT /note= "Asn in native MSP-1-42 (N262Q mutation)"
XX PN WO9920766-A2.
XX PD 29-APR-1999.
XX PF 20-OCT-1998; 98WO-US222225.
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 97US-0062592.
XX PA (GENZ ) GENZYME TRANSGENICS CORP.
XX PI Chen LH, Meade H;
XX DR WPI; 1999-302742/25.
XX DR N-PSDB; AAX25593.
XX PT New modified recombinant nucleic acid sequences useful for producing
XX PT malarial DNA vaccine
XX PS Example; Fig 11; 43pp; English.
XX CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The sequence has been modified to include
CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
CC tag. In addition, N181Q and N262Q mutations have been introduced to
CC eliminate 2 N-glycosylation sites. These modifications allow the
CC MSP-1-42 protein to be expressed in the milk of transgenic mice.
CC The invention generally relates to modified recombinant nucleic
CC acid sequences and methods for increasing the mRNA levels and
CC protein expression of proteins that are difficult to express in
CC cell culture systems, mammalian cell culture systems or in
CC transgenic animals. The preferred difficult protein candidates for
CC expression are those derived from lower organisms such as parasites,
CC bacteria and viruses that have DNA coding sequences of high AT
CC content or which have mRNA instability motifs or rare codons
CC relative to the recombinant expression system to be used.
XX CC
XX SQ Sequence 376 AA;

Query Match 57.4%; Score 1184.5; DB 20; Length 376;
Best Local Similarity 59.9%; Pred. No. 8.1e-76;
Matches 221; Conservative 53; Mismatches 74; Indels 21; Gaps 2;

QY 6 MDNLGFEYEDVYIKPLAGVYRSLLKQIEKNIFTNLDILNSRLKRRKRYFDVL 65
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
22 idnllskieneyevlykplagvyrslkqlennvmtfnvkvdkilnsrlnkrenfknvl 81

QY 66 ESDLMOFKHSSNEYIIEDSFKLNSQKNTLLKSKYIKESVENDIKFAQEGISYVEKV 125
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
82 esdlipkydltssnyvkvdkpkyfinkkrdkfissynpykdsidtdinfandvgyykil 141

QY 126 LAKYKDDLESIKKVIKEEKEKFPSPPTTPPSPAKTDEQKKESKFLPFLTNIETLYNNLV 185

```


CC This is the amino acid sequence of a recombinant protein comprising
 CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
 CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
 CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
 CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX Sequence 116 AA;

Query Match 31.0%; Score 640; DB 18; Length 116;
 Best Local Similarity 100.0%; Pred. No. 4.2e-38;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 NISQHCVCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCDADA 340
 DB 3 nlsqhcvcvkqcpnsgcfrrhldeereckcllnkqegdkcvenpnptcnennggcdada 62
 QY 341 KCTEEDSGSGKKTCTCKTPDSYPLFDGIFCSSNFIISFLILMLILYSFI 394
 DB 63 kcteedsgsgkktctcktpdpsylfdgicssnfigisflilmlilysfi 116

RESULT 12
 AAB37609
 ID AAB37609 standard; Protein; 108 AA.

AC AAB37609;
 XX 27-FEB-2001 (first entry)
 XX Merozoite surface protein-119.

XX Merozoite surface protein; protazoacide; vaccine; malaria.
 XX Plasmodium falciparum.
 XX WO200063245-A2.

XX 26-OCT-2000.
 XX 20-APR-2000; 2000WO-GB01558.
 XX 20-APR-1999; 99GB-0009072.
 XX 13-MAY-1999; 99US-0311817.
 XX 25-MAY-1999; 99CA-2271451.

XX (MEDI-) MEDICAL RES COUNCIL.
 XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
 XX WPI; 2001-015762/02.

XX N-PSDB; AAC68977.
 XX Novel variants of the C-terminal fragment of Plasmodium merozoite
 XX surface protein-1, useful as vaccines for treating or preventing
 XX malaria -

XX Example 5; Fig 15; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal
 XX fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 XX non-natural variants have reduced affinity for at least 1 antibody
 XX capable of blocking a second antibody that inhibits the proteolytic
 XX cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
 XX one third antibody that inhibits the proteolytic cleavage of Plasmodium
 XX MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
 XX present invention are useful for immunising a mammal against malaria, and
 XX can be used to treat malaria. The present sequence is MSP-119 protein.

XX Sequence 108 AA;

Query Match 26.9%; Score 554; DB 22; Length 108;
 Best Local Similarity 92.4%; Pred. No. 4.5e-32;
 Matches 97; Conservative 3; Mismatches 1; Indels 4; Gaps 1;

QY 272 IEQFQDMLNISQHCVCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNTCNE 331
 DB 8 ieqrh----niaqhcvkqcpnsgcfrrhldeereckcllnkqegdkcvenpnptcne 63
 QY 332 NNGCDADAKCTEEDSGSGKKTCTCKTPDSYPLFDGIFCSSN 376
 DB 64 nnggcdadakcteedsgsgkktctcktpdpsylfdgicssn 108

RESULT 13
 AAB37608
 ID AAB37608 standard; Protein; 96 AA.

AC AAB37608;
 XX 27-FEB-2001 (first entry)
 XX Merozoite surface protein-1.

XX Merozoite surface protein; protazoacide; vaccine; malaria.
 XX Plasmodium falciparum.
 XX WO200063245-A2.

XX 26-OCT-2000.
 XX 20-APR-2000; 2000WO-GB01558.
 XX 20-APR-1999; 99GB-0009072.
 XX 13-MAY-1999; 99US-0311817.
 XX 25-MAY-1999; 99CA-2271451.

XX (MEDI-) MEDICAL RES COUNCIL.
 XX Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
 XX WPI; 2001-015762/02.

XX Novel variants of the C-terminal fragment of Plasmodium merozoite
 XX surface protein-1, useful as vaccines for treating or preventing
 XX malaria -
 XX Example 2; Page 48; 126pp; English.

XX The present invention relates to non-natural variants of a C-terminal
 XX fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 XX non-natural variants have reduced affinity for at least 1 antibody
 XX capable of blocking a second antibody that inhibits the proteolytic
 XX cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
 XX one third antibody that inhibits the proteolytic cleavage of Plasmodium
 XX MSP-1.4.2, compared to natural MSP-1.1.9. The present sequence is the
 XX wild-type MSP-1 protein. This sequence was used to generate the variants
 XX of the present invention. The non-natural variants of the present
 XX invention are useful for immunising a mammal against malaria, and can be
 XX used to treat malaria.

XX Sequence 96 AA;

Query Match 26.8%; Score 553; DB 22; Length 96;
 Best Local Similarity 99.0%; Pred. No. 4.5e-32;
 Matches 95; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 281 NISQHCVCVKKQCPNSGCFRHLDERECKCLLNKQEGDKCVENPNTCNENGGCDADA 340
 DB 1 nlsqhcvcvkqcpnsgcfrrhldeereckcllnkqegdkcvenpnptcnennggcdada 60

Matches	94;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0
QY	341	KCTEEDSGSGKKITCECTKPSYPLFDGIFCSSN	376						
Db	61	ktceedsgsgkkitcectkpsypifdgifcssn	96						
RESULT	14								
ID	AAW22593	standard; Protein; 127 AA.							
XX	AC	AAW22593;							
XX	AC	AAW22593;							
DT	25-MAR-1998	(first entry)							
XX	DE	PfMSP1(p19)S protein sequence.							
XX	Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;								
XX	Plasmodium falciparum; malaria; vaccine; immunity; epitope.								
XX	Chimeric - Plasmodium vivax.								
XX	Chimeric - Plasmodium falciparum.								
XX	Key	Location/Qualifiers							
FT	Peptide	1..19							
FT	Protein	/note= "signal peptide"							
FT	Region	20..127							
FT	Region	/note= "mature protein"							
FT	Region	1..32							
FT	Region	/note= "derived from P. vivax MSP1"							
FT	Region	33..34							
FT	Region	/note= "encoded by restriction enzyme sequence used to create the chimeric sequence"							
FT	Region	35..127							
FT	Region	/note= "derived from P. falciparum C-terminal p19 fragment of MSP1"							
XX	WO9730159-A2.								
XX	21-AUG-1997.								
XX	14-FEB-1997;	97WO-FR00291.							
XX	14-FEB-1996;	96FR-0001821.							
XX	(INSP) INST PASTEUR.								
XX	(UYN) UNIV NEW YORK STATE.								
XX	Barnwell JW, Longacre-Andre S, Mendis K, Nato F;								
XX	Roth C;								
XX	WPI; 1997-425034/39.								
XX	P-PSDB; AAW22592.								
XX	Recombinant protein containing Plasmodium merozoite surface								
XX	protein-1 p42 fragment - useful in antimalarial vaccines, also new								
XX	antibodies for diagnosis and protein purification								
XX	Disclosure; Fig 1C; 85pp; French.								
XX	This is the amino acid sequence of a chimeric protein comprising amino								
XX	acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)								
XX	linked to amino acids Asn163-Ser1705 which correspond to the 19 kD								
XX	C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.								
XX	p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.								
XX	The recombinant protein can be used for the production of anti-malarial								
XX	vaccines, where the p19 fragment provides a high level of protective								
XX	immunity since it includes epitopes not presented in the p42 fragment.								
XX	Sequence 127 AA;								
Query Match	26.5%;	Score 546;	DB 18;	Length 127;					
Best Local Similarity	97.9%;	Pred. NO. 2e-31;							

Query Match 26.5%; Score 546; DB 18; Length 127;
Best Local Similarity 97.9%; Pred. No. 2e-31;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 43.63 Seconds
(without alignments)
687.893 Million cell updates/sec

Title: US-09-500-376-2

Perfect score: 2062

Sequence: 1 AISTVMDNILSGFENEYDVI.....SNFLGIFLLMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 7617452 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2062	100.0	1726	1 SAZ08M	major merozoite su
2	2043.5	99.1	651	2 S47282	merozoite surface
3	2043	99.1	1726	2 A45948	major merozoite su
4	2027	98.3	1701	2 A54498	major merozoite su
5	2006	97.3	1701	2 A26968	major merozoite su
6	1289.5	62.5	400	2 A45545	major merozoite su
7	1289.5	62.5	1639	2 S05603	major merozoite su
8	1166	56.5	1631	1 SAZ0K1	major merozoite su
9	1124	54.5	1640	2 A24594	probable major sur
10	843	40.9	1726	2 A39401	merozoite surface
11	832	40.3	1751	2 A45604	major blood-stage
12	613.5	29.8	680	2 A28121	major merozoite su
13	613.5	29.8	1772	2 A45532	major merozoite su
14	583	28.3	1785	2 A45546	major merozoite su
15	180.5	8.8	980	2 E71606	hypothetical prote
16	172	8.3	2166	2 G70163	hypothetical prote
17	162.5	7.9	1939	2 T18372	repeat organellar
18	159.5	7.7	1169	2 A64505	pl15 homolog - Met
19	156	7.6	2401	2 T28676	rhostry protein -
20	155.5	7.5	1619	2 T18499	hypothetical prote
21	153	7.4	1130	2 T34081	hypothetical prote
22	152.5	7.4	2269	2 T28677	rhostry protein -
23	150	7.3	1156	2 B70356	chromosome assembl
24	145	7.0	3394	2 T18501	hypothetical prote
25	144.5	7.0	1086	2 S16752	major merozoite su
26	143	6.9	652	2 B59102	hypothetical prote
27	142	6.9	1250	2 E81339	probable restricti
28	141	6.8	442	2 T18507	hypothetical prote
29	141	6.8	1005	2 A64465	hypothetical prote

30 141 6.8 1127 2 T28317
31 141 6.8 1187 2 T18355
32 140 6.8 624 2 PC6003
33 139.5 6.8 671 2 H64502
34 139.5 6.8 1225 2 A49464
35 138.5 6.7 1979 2 C71622
36 138 6.7 1079 2 T18356
37 138 6.7 1313 2 F96673
38 137.5 6.7 819 2 E70105
39 137 6.6 2116 2 A26655
40 136.5 6.6 2829 2 A42771
41 136 6.6 1679 2 S48385
42 136 6.6 2573 2 D71614
43 135 6.5 821 2 S67087
44 135 6.5 880 2 F75103
45 135 6.5 1228 2 A57384

ALIGNMENTS

RESULT 1

SAZ08M

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: 195K glycoprotein
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361

R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.

Nucleic Acids Res. 14, 3311-3323, 1986

A:Title: Variation in the gene encoding a major merozoite surface antigen of the huma

A:Reference number: A23386; MUID:86205236

A:Accession: A23386

A:Molecule type: DNA

A:Residues: 1-1104 <WEB1>

A:Cross-references: EMBL:X03831

R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.

Nucleic Acids Res. 16, 1206, 1988

A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria

A:Reference number: S06361; MUID:88143999

A:Accession: S06361

A:Molecule type: DNA

A:Residues: 1104-1726 <WEB2>

A:Cross-references: EMBL:X03831

C:Comment: The merozoite stages of different strains have strain-specific surface ant

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1726/Product: major merozoite surface antigen #status predicted <MAT>

F;67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)

F;757-765/Region: 3-residue repeats (T-E-E)

F;133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carb

Query Match 100.0%; Score 2062; DB 1; Length 1726;

Best Local Similarity 100.0%; Pred. No. 2.1e-100;

Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AISTVMDNILSGFENEYDVIYLRPLAGVYSLKQIKENITFTFNLDILNSLRKKRY 60

|||||

Db 1333 AISTVMDNILSGFENEYDVIYLRPLAGVYSLKQIKENITFTFNLDILNSLRKKRY 1392

|||||

QY 61 FLDVLESDLMQFKHISNEYIIDSFKLLNSEQNTLLSKYIKESVENDIKFAQSGIS 120

|||||

Db 1393 FLDVLESDLMQFKHISNEYIIDSFKLLNSEQNTLLSKYIKESVENDIKFAQSGIS 1452

|||||

QY 121 YEKVLAKYKDDLESIKVKIKEEKFPSPPTPPAKTDEOKKESKFLPFTNIETL 180

|||||

Db 1453 YEKVLAKYKDDLESIKVKIKEEKFPSPPTPPAKTDEOKKESKFLPFTNIETL 1512

|||||

QY 181 YNNLVNKKIDDYILNLKAKINDCINVEKDEAHVTKLSDLKAIDDKIDLKFNHNDFEAIKK 240

|||||

Db 1513 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKTITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572

QY 241 LINDTAKMDGLKLLSTGLVQNPNTIISKLEGGKFDMDLNIHQHCVKQKCPENSGCFR 300
|||||
Db 1573 LINDTAKMDGLKLLSTGLVQNPNTIISKLEGGKFDMDLNIHQHCVKQKCPENSGCFR 1632
|||||

QY 301 HLDRECKCLLNYKQBGDKCVENPNPTCNENNGGCCDADAKTEEDSGNGKKITCCECTK 360
|||||

Db 1633 HLDRECKCLLNYKQBGDKCVENPNPTCNENNGGCCDADAKTEEDSGNGKKITCCECTK 1692
|||||

QY 361 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 394
|||||

Db 1693 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 1726
|||||

RESULT 2
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71) (H
C:Species: Plasmodium falciparum
A:Variety: strain RO-71
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C:Accession: S47282
R:Jolley, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A:Description: Plasmodium falciparum: recombination within the C-terminal region of mer
A:Reference number: S47282
A:Accession: S47282
A:Molecule type: DNA
A:Residues: 1-651 <TOL>
A:Cross-references: EMBL:235329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A:Experimental source: strain RO-71
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 99.1%; Score 2043.5; DB 2; Length 651;
Best Local Similarity 99.5%; Pred. No. 6.6e-100;
Matches 392; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNFTNLNLDILNSRLKRRKY 60
|||||

Db 259 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNFTNLNLDILNSRLKRRKY 318
|||||

QY 61 FLDVLESDDLMOQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
|||||

Db 319 FLDVLESDDLMOQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 378
|||||

QY 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
|||||

Db 379 YEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 437
|||||

QY 181 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKTITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
|||||

Db 438 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKTITKLSDLKAIDDKIDLFKNHNDFEAIKK 497
|||||

QY 241 LINDTAKMDGLKLLSTGLVQNPNTIISKLEGGKFDMDLNIHQHCVKQKCPENSGCFR 300
|||||

Db 498 LINDTAKMDGLKLLSTGLVQNPNTIISKLEGGKFDMDLNIHQHCVKQKCPENSGCFR 557
|||||

QY 301 HLDRECKCLLNYKQBGDKCVENPNPTCNENNGGCCDADAKTEEDSGNGKKITCCECTK 360
|||||

Db 558 HLDRECKCLLNYKQBGDKCVENPNPTCNENNGGCCDADAKTEEDSGNGKKITCCECTK 617
|||||

QY 361 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 394
|||||

Db 618 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 651
|||||

RESULT 3
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000

C:Accession: A45948
R:Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major me
A:Reference number: A45948; MUID:89005525
A:Accession: A45948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1726 <CHA>
A:Cross-references: GB:M37213
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 99.1%; Score 2043; DB 2; Length 1726;
Best Local Similarity 99.2%; Pred. No. 2e-99; Mismatches 0; Indels 0; Gaps 0;
Matches 391; Conservative 0

QY 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNFTNLNLDILNSRLKRRKY 60
|||||

Db 1333 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNFTNLNLDILNSRLKRRKY 1392
|||||

QY 61 FLDVLESDDLMOQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
|||||

Db 1393 FLDVLESDDLMOQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 1452
|||||

QY 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
|||||

Db 1453 YEKVLAKYKDDLESIKKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 1512
|||||

QY 181 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKTITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
|||||

Db 1513 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKTITKLSDLKAIDDKIDLFKNHNDFEAIKK 1572
|||||

QY 241 LINDTAKMDGLKLLSTGLVQNPNTIISKLEGGKFDMDLNIHQHCVKQKCPENSGCFR 300
|||||

Db 1573 LINDTAKMDGLKLLSTGLVQNPNTIISKLEGGKFDMDLNIHQHCVKQKCPENSGCFR 1632
|||||

QY 301 HLDRECKCLLNYKQBGDKCVENPNPTCNENNGGCCDADAKTEEDSGNGKKITCCECTK 360
|||||

Db 1633 HLDRECKCLLNYKQBGDKCVENPNPTCNENNGGCCDADAKTEEDSGNGKKITCCECTK 1692
|||||

QY 361 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 394
|||||

Db 1693 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 1726
|||||

RESULT 4
A54498
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasm
A:Reference number: A54498; MUID:88142999
A:Accession: A54498
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PET>
A:Cross-references: GB:M19143; NID:g160412; PIDN:AAA29653.1; PID:g160413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 98.3%; Score 2027; DB 2; Length 1701;
Best Local Similarity 98.7%; Pred. No. 1.4e-98;
Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNFTNLNLDILNSRLKRRKY 60
|||||

Db 1308 AISVTMDNILSGFENEYDVYILKPLAGVYRSLLKQIEKNFTNLNLDILNSRLKRRKY 1367
|||||

QY 61 FLDVLESLDMQFKHSSNEYIIEFSLKLNSEQNTLLKSKYIKESVENDIKFAQEGIS 120
|||||
Db 1368 FLDVLESLDMQFKHSSNEYIIEFSLKLNSEQNTLLKSKYIKESVENDIKFAQEGIS 1427
|||||
QY 121 YIEKVLAKYKDDLESIRKVKKEKEKFPSPPTPPSPAKTDEOKKESKFLPFLTNIEETL 180
|||||
Db 1428 YIEKVLAKYKDDLESIRKVKKEKEKFPSPPTPPSPAKTDEOKKESKFLPFLTNIEETL 1487
|||||
QY 181 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKFNHNDFAIKK 240
|||||
Db 1488 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKFNHNDFAIKK 1547
|||||
QY 241 LINDDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 300
|||||
Db 1548 LINDDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 1607
|||||
QY 301 HLDRECKCLLNYKQEGDKCVENPNTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
|||||
Db 1608 HLDRECKCLLNYKQEGDKCVENPNTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1667
|||||
QY 361 PDSYPLFDGIFCSCSSNFIKLSLILMLILYSFI 394
|||||
Db 1668 PDSYPLFDGIFCSCSSNFIKLSLILMLILYSFI 1701
|||||
RESULT 5
A26868
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C/Species: Plasmodium falciparum
C/Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000
C/Accession: A26868
R:Tanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.
J. Mol. Biol. 195, 273-287, 1987
A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>
Query Match 97.3%; Score 2006; DB 2; Length 1701;
Best Local Similarity 98.0%; Pred. No. 1.7e-97;
Matches 386; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 1 AISVTMDNILSGFENEYDVIYKPLAGVYRSKKQIEKNITFTNLDILNSRLKRRKY 60
|||||
Db 1308 AISVTMDNILSGFENEYDVIYKPLAGVYRSKKQIEKNITFTNLDILNSRLKRRKY 1367
|||||
QY 61 FLDVLESLDMQFKHSSNEYIIEFSLKLNSEQNTLLKSKYIKESVENDIKFAQEGIS 120
|||||
Db 1368 FLDVLESLDMQFKHSSNEYIIEFSLKLNSEQNTLLKSKYIKESVENDIKFAQEGIS 1427
|||||
QY 121 YIEKVLAKYKDDLESIRKVKKEKEKFPSPPTPPSPAKTDEOKKESKFLPFLTNIEETL 180
|||||
Db 1428 YIEKVLAKYKDDLESIRKVKKEKEKFPSPPTPPSPAKTDEOKKESKFLPFLTNIEETL 1487
|||||
QY 181 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKFNHNDFAIKK 240
|||||
Db 1488 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKFNHNDFAIKK 1547
|||||
QY 241 LINDDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 300
|||||
Db 1548 LINDDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 1607
|||||
QY 301 HLDRECKCLLNYKQEGDKCVENPNTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 360
|||||
Db 1608 HLDRECKCLLNYKQEGDKCVENPNTCNENNGGCDADAKCTEEDSGSNGKKITCECTK 1667
|||||

QY 361 PDSYPLFDGIFCSCSSNFIKLSLILMLILYSFI 394
|||||
Db 1668 PDSYPLFDGIFCSCSSNFIKLSLILMLILYSFI 1701
|||||
RESULT 6
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments
C/Species: Plasmodium falciparum
C/Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C/Accession: A45545
R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protei
A:Reference number: A45545; MUID:92131048
A:Accession: A45545
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-400 <BLA>
A>Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIIP:77621)
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
Query Match 62.5%; Score 1289.5; DB 2; Length 400;
Best Local Similarity 62.3%; Pred. No. 1e-60;
Matches 243; Conservative 54; Mismatches 72; Indels 21; Gaps 2;
QY 6 MONTLSGFENEYDVIYKPLAGVYRSKKQIEKNITFTNLDILNSRLKRRKYFLDVL 65
|||||
Db 31 IDNLTSLKIENEYEVLYLPLAGVYRSKKQIEKNITFTNLDILNSRLKRRKYFLDVL 90
|||||
QY 66 ESDLMOQFKHSSNEYIIEFSLKLNSEQNTLLKSKYIKESVENDIKFAQEGISYVEKV 125
|||||
Db 91 ESDLIPYKDTSSVYVYKPLAGVYRSKKQIEKNITFTNLDILNSRLKRRKYFLDVL 150
|||||
QY 126 LAKYKDDLESIRKVKKEKEKFPSPPTPPSPAKTDEOKKESKFLPFLTNIEETLYNNLV 185
|||||
Db 151 SEKYKSLDLSIKKYI-----NDKQGENEKYLPFLNNIEETLYKTVN 190
|||||
QY 186 NKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKFNHNDFAIKKLLND 245
|||||
Db 191 DKIDLFVHLEAKVLYNTYKSNVEVEIKELNYLKTQDKLADFKNNNFVGIADLSTDY 250
|||||
QY 246 TKDMLCKLLSTGLV-QNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFRHLDE 304
|||||
Db 251 NNNLLKFLSTGVFVFNLAKTVLSNLDLQMLNLSQHCVKKQCPNSGCFRHLDE 310
|||||
QY 305 REECKCLLNYKQEGDKCVENPNTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDY 364
|||||
Db 311 REECKCLLNYKQEGDKCVENPNTCNENNGGCDADAKCTEEDSGSNGKKITCECTKPDY 370
|||||
QY 365 PLFDGIFCSCSSNFIKLSLILMLILYSFI 394
|||||
Db 371 PLFDGIFCSCSSNFIKLSLILMLILYSFI 400
|||||
RESULT 7
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C/Species: Plasmodium falciparum
C/Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C/Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R:Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989

A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasmodium falciparum
A;Reference number: S04850; MUID:89345116

A;Accession: S04850
A;Molecule type: mRNA
A;Residues: 1504-1639 <MYL2>
A;Cross-references: EMBL:X15063
C;Superfamily: major merozoite surface antigen
C;Keywords: glycoprotein; merozoite; surface antigen
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match	62.5%;	Score 1289.5;	DB 2;	Length 1639;
Best Local Similarity	62.3%;	Pred. No. 4.9e-60;		
Matches 243:	Conservative	54;	Mismatches 72;	Indels 21; Gaps 2;

Qy	6	MDNILSGFENEYDVLYKPLAGVYSRLKKOIEKNFTFNLNLDIINSRLKRRKYFLDVL	65
		: : : : : : : : :	
Db	1270	IDNLSKIENYEYVLYKPLAGVYSRLKKOLENNVTNFVNWKDIINSRFKRENFKNVL	1329

Qy	66	ESDLMQFHHISSNEYIIEDSFLLNSEQNLTLLSKYKIKESVENDIKFAEGISYYEKV	125
		: : : : : : : : :	
Db	1330	ESDLIPKDLTSSNYVVKDYPKFNKEKRDKFLSSNYIKDSIDTDINFANDVLGYKIL	1389

```
Qy 126 LAKYKDDJESIKKVIKEKEKFPSPPTTTPPSAKTDEQKESKFPLFTNIETLYNNLV 195
      ||| ||:|||| |
      |:| :|:|||| |
Db 1390 SEKYKSDJDSIKYI-----NDKOGNEKYLPFLNNIETLYKTVN 1429
```

Qy	186	NKIDYILNKA KINCNDCKVEKDEAHVKITKLSDLKAIDDKIDLFKHNDFEAIKKLINDD	245
	:	: : :	:
Db	1430	DKIDLFTVHLEAKLYNTYEKSNSVEVKIKELNYLTODKIADFEKNNNFVGTLSDTY	1489

Qy 246 TKKDMGLKLLSTGLV-QNEPNTIISKLECKFQDMLNISQHCVKKQCPENSGCFRHLDE 304
:::||||:| |::|::| |||||||:|||||
Dβ 1490 NHNNLTKFLSTGVAFFENIAKTVLNLDIGNIOGMNISOHCVKKQCPONSCFRHLDE 1549

Qy 305 REECKCLLNYYKQEGDKCVENPNTCNENNCGCDAADAKCTEEDSGSNKKKITCECTKPDSY 364
|||||
|||||
Db 1550 REECKCLLNYYKQEGDKCVENPNTCNENNCGCDAADAKCTEEDSGSNKKKITCECTKPDSY 1609
|||||
|||||

Qy 365 PLFDGIFCCSSNFIIGISFLLILMLILYSFI 394
|||||
Db 1610 PLFDGIFCCSSNFIIGISFLLILMLILYSFI 1639

8
RESULT
SAZQK1

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (sub
C;Species: Plasmodium falciparum
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 09-Jun-2000
C;Accession: A25120

EMBO J. 4, 3823-3829, 1985
A:Title: Polymorphism of the precursor for the major surface antigens of plasmodium falciparum
A:Reference number: A91030; PMID:86136024
A:Author(s): Goman, M.; Bore, N.; Hyde, J.B.; Scarle, J.; Cella, G.; Stannenberg, H.; Mackay, W.

A: Accession: [AZJ120](#)
 A: Molecule type: DNA
 A: Residues: 1-1631 <MAC>
 C: Comment: The merozoitic stages of different strains have strain-specific surface antigenic proteins.

C:Superfamily: major merozoite surface antigen
C:Comment: P. taichiparum has three stages: sporozoite, merozoite, and gametocyte. The
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote
F:1-19/Domain: signal sequence #status predicted <sig>

F:20-163n/Product:	major microzoelle surface
F:67-84/Region:	3-residue repeats (\$G-T/P)
F:1614-1631/Domains:	membrane anchor #status <MBN>
F:97-259/755-759-835:	911-955-1049-1156-1165-1436-1563/Binding site: carbonylate (Asn)

Query Match	56.5%	Score 1166;	DB 1;	Length 1631;
Best Local Similarity	58.1%	Pred. No. 1	46-53;	

Matches	227	Conservative	57	Mismatches	85	Indels	42	Gaps	3
---------	-----	--------------	----	------------	----	--------	----	------	---

Qy	6	MDNLTSGFENEYDVLYLKPLAGVYRSLKKQIEKNITFTNLDNLNSRLKRRKYFDVL	65
Db	1261	IDNLTSIENEYEVLYLKPLAGVYRSLKKQLENNYMTFNVNYKILNSRFRNKENFNVL	1320
Qy	66	ESDLMQFKHLSSENYIIESDFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGISYEKV	125
		: : : : : : : : : : : :	
Db	1321	ESDLIPYKDLTSSNYVYVQPKPLNKRKDFKLLSSVNYIKSDIDTFINFADVLYGYKIL	1380
Qy	126	LAKYKDDLESIKKVIKEEKEKFPSSPPTTPPSPAKTDQEKRESKLPPLTNIETLYNNLV	185
Db	1381	SEKYKSDLSIDIKKYI-----NDKQGENEKLVPFLNIIETLYKTVN	1420

DQ 186 NKIDDYLNLKAKINDCNVEKDFAHVKITLSLDKAIDDKIDLFKHNDFEAIKKLINDD 245
 :|||:::||::: || |::| |::| ::|::| |::| |::| |
Db 1421 DKDLFVHLEAVLNYYTEKSNEVKIKELIYKTIQDKLADEFKANNFVCIALSTDY 1480

Qy 246 TKDMLGKLLSTGLV-QNFPNTIISKLIEGKF-QDMLNISQHCVKKQCPENSGCFRHL D 303
 : : | ||||| : : | : | : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 1481 NHNNLLTFELSTGMVFENLKLSILSNLDWKLARVYKHFTPMRKKTMTIQNSGCFRHL D 1540

[illegible]

QY 364 YPLFDGIFCSSNFGISFLLIMLILYSFI 394
 |||||
 Db 1601 YPLFDGIFCSSNFGISFLLIMLILYSFI 1631

RESULT 9
A24594

C:Species: *Plasmodium falciparum*
C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jun-2000
C:Accession: A24594

A:Title: Primary structure of the precursor to the three major surface anti-
A:Reference number: A24594; MUID:86014355

A: Molecule type: DNA
A: Residues: 1-1640 <HOL>
C: Superfamily: major merozoite surface antigen

Query Match	54.5%	Score 1124;	DB 2;	Length 1640;
Best Local Similarity	56.3%	Pred. No. 2.2e-51;		
Matches 220; Conservative	60;	Mismatches 89;	Indels 22;	Gaps 3;

Qy 6 MDNLSGFENEYDYIYLKPLAGVVSRLKKOIEKNITFTNLNDILNSRLKRKYFDVL 65
 :||||| :||| |:| | ||| | | | | | | | | | | | | | | | | | | | | | |
Dz IDNLSKIENYEYLYLKLPLAGVVSRLLKKOLENNVTFFNVVKDILNSRFKRENFANVL 1329

QY	66	ESDLMQKXHISSNEYIIEDSFKLINSEQNTLLSKYKIRESVENDIKFAQGSIYYEKV	125
		: : : : : : : : : : : : : : : :	
Db	1330	ESDLPYKDLTSSNYVKDPYKFINKERDKFLSSNYYIKDSIDTDINFANDVLGYKKIL	1389

QY 126 LAKYKDDLESIKKVIKEEXEKFPSPPPTTTPSPAKTDEQKESKFPLFTNTIETLYNNLV 185
 III III:III I :I: I:III IIIII ;
 Db 1390 SERYSKSDLSIKKVI-----NDKOGNEKYLPFLNNIETLYKTVN 1429

QY 186 NKIDDYLINLAKAKINDCNVEKDEAHVKITKLSDLKAIDDKDIDLFKNHNDFEAIKKLINDD 245
:III::I:I::: : III::I:I::I:I::I:I::I:I::I:I::I:I::I:I::I:I::I:I::I:I::

Dd 1430 DKIDLFLVHLEAKLYNVTYEKSNNVEVKIKELNYLTIODKLADFKKNNNFVGIAIDSTDY 1489

QY 246 TKKDMGLKLLSTGLV-QNFPNTIISKLIEKF-QDMLNISQHVKKQCPENSGCFRHL 303
:::| ||||:| ::| |:::| : : : | ::||| |||
Db 1490 NHHNIIYFIISGVVFENILKSVISNLLDKMLARYVYKHFTTPMRKKTMIOSSGCFRHL 1549

QY 304 ERECKCILLNYKQEGDKCVENPPTCNNGGCDADAKCTEEDSGSNGKKITCECTAPDS 363
|||||
db 1550 ERECKCIIINYKORGSKVCNSNPPTCNNGGCDADAKCTEEDSGSNGKKITCECTAPDC 1609

A:Cross-references: GB:AE001417; GB:AE001362; MID:g3845271; PIDN:AAC71942.1; PID:g384527
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0765w
C:Superfamily: hypothetical protein MJ1322

```
Query Match      8.8%; Score 180.5; DB 2; Length 980;
Best Local Similarity 24.0%; Pred. No. 0.028;
Matches 103; Conservative 75; Mismatches 141; Indels 111; Gaps 23;

Qy 14 ENEYDVIIYLRPLAGVYRSLLKQIEKNI-----FTFNLNLDIL---NSRLKK 57
Db 248 QNYI---YLKK---EYDLKNKELEKNIHGKLEHLSHCYEEENQKLNNEETKRNSPIKN 301
Qy 58 RKYFLDVL---ESLMOFKHIS-----SNEVIIED-----SFK 87
Db 302 KDRKIDLLTNIENELLKKKEINNKLMEKQNVKNNQOLLKIDKIDENEKMNHVNKLQNE 361
Qy 88 LLNSEQKNTLL-KSYKYIKESVENDIRFAQEGISYIEKVLAKYKDDLESIKK----- 138
Db 362 LKRELQNKICISKIEFCCKEKEDKIKNLEDDLLEKKKCIENLKDDELINIKKKMEDKMHM 421
Qy 139 -----VIKEEKEKPPSPPTTPPSPAKTDEQ-----KKESKLPFLFTNIETLYN-NLVNKI 188
Db 422 TNEMDLLSNKYVEELNRINKTYEKNIVELNNELDVIRKKLNDEEFLKEEERKKNIDMYKI 481
Qy 189 DDYLINLKAKINDC-NVEKDSAHVKITKLSDLKAIDDKIDLFKNHNDPEAKKLINDDTK 247
Db 482 KEYEIQIKERENEIDSLKKNQNLHVLNKEEL---NEKEITLKNKYDKE-INMIIEQYNK 537
Qy 248 -----KDWLGLKLLSTGLVQNPNTIISKLEIGKQF-----DMLNISQHCQVKQCPE 294
Db 538 KIQEKMNLNNKIKS-MDQTHKNQIEEQEENKELKRLKNVCDM-NLQSOILIK-----E 591
Qy 295 NSGCFRHLDER-ECKCLLNKY-----QEGDKCVENPNPTCNENNGGCDADAKCTEE 345
Db 592 NE---KHMQERKVEYKNNLLKQKQDELKNIIOEYDERIEIQNKEMEDIVNDCCEKLK---- 644
Qy 346 DSGSGNGKKIT 355
Db 645 QAKINNKKLT 654
```

Search completed: August 8, 2001, 12:30:49
Job time: 97 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 24.79 Seconds
(without alignments)
544.440 Million cell updates/sec

Title: US-09-500-376-2

Perfect score: 2062

Sequence: 1 AISVTMDNITLGFENEYDVI.....SNFLGISFLILMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2062	100.0	1726	1 MSPI_PLAFC	P04934 plasmodium
2	2059	99.9	1726	1 MSPI_PLAFC	P50495 plasmodium
3	2027	98.3	1701	1 MSPI_PLAFC	P13819 plasmodium
4	2027	98.3	1701	1 MSPI_PLAFC	P08569 plasmodium
5	2025.5	98.2	1682	1 MSPI_PLAFC	P19598 plasmodium
6	1289.5	62.5	1630	1 MSPI_PLAFC	P04932 plasmodium
7	1289.5	62.5	1639	1 MSPI_PLAFC	P04933 plasmodium
8	613.5	29.8	1772	1 MSPI_PLAFC	P13828 plasmodium
9	153	7.4	1130	1 YL17_CAEEL	Q11102 caenorhabdi
10	139.5	6.8	1225	1 SMC1_YEAST	P32908 saccharomyc
11	137	6.6	2116	1 MYS2_DICDI	P08799 dictyosteli
12	136.5	6.6	2869	1 RBPI_PLAFC	P00798 plasmodium
13	136	6.6	1679	1 Y109_YEAST	P04057 saccharomyc
14	135	6.5	1228	1 ECM_HUMAN	Q13201 homo sapien
15	130.5	6.3	1790	1 USO1_YEAST	P25386 saccharomyc
16	129	6.3	944	1 NUF1_YEAST	P32380 saccharomyc
17	129	6.3	2245	1 MYSJ_DICDI	P54697 dictyosteli
18	128	6.2	796	1 YFC3_YEAST	P43573 saccharomyc
19	128	6.2	1038	1 CIN8_YEAST	P27895 saccharomyc
20	128	6.2	1251	1 RBP2_PLAFC	Q00799 plasmodium
21	127	6.2	540	1 YK26_YEAST	P36112 saccharomyc
22	127	6.2	729	1 KAR3_YEAST	P17119 saccharomyc
23	127	6.2	1324	1 CUT3_SCHPO	P41004 schizosacch
24	126	6.1	1312	1 RA50_YEAST	P12753 saccharomyc
25	125.5	6.1	622	1 YAE7_YEAST	P39723 saccharomyc
26	125	6.1	471	1 SG3_MOUSE	P47867 mus musculu
27	124.5	6.0	1875	1 MLPI_YEAST	Q02455 saccharomyc
28	124	6.0	1928	1 MYS1_YEAST	P08964 saccharomyc
29	123.5	6.0	703	1 HS83_PHANI	P51819 pharbitis n
30	123.5	6.0	798	1 ITB1_HUMAN	P05556 homo sapien
31	122.5	5.9	1230	1 SMC3_YEAST	P47037 saccharomyc
32	122	5.9	399	1 RMS3_PENUR	P47907 penicillium
33	122	5.9	1170	1 SMC2_YEAST	P38989 saccharomyc

RESULT 1

ID	MSPI_PLAFC	STANDARD;	PRT;	1726 AA.
AC	P04934;			
DT	13-AUG-1987 (Rel. 05, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	MEEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEEROZOITE SURFACE ANTIGENS) (PMMSA) (P195).			
DE	GN MSP-1.			
OS	Plasmodium falciparum (isolate Camp / Malaysia).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5835;			
RN	[1]			
RP	SEQUENCE OF 1-1103 FROM N.A.			
RX	MEDLINE=86205236; PubMed=3517809;			
RA	Weber J.L., Leininger W.M., Lyon J.A.;			
RT	"Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite Plasmodium falciparum."			
RL	Nucleic Acids Res. 16:1206-1206(1988).			
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).			
CC	-1- PTM: MEEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X03831; CAA27446.1; -			
DR	PIR: A23386; SAZQGM.			
DR	InterPro: IPR000561; -			
DR	Pfam: PF00008; EGF; 1.			
KW	Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;			
KW	Transmembrane; GPI-anchor.			
FT	SIGNAL 1 19 POTENTIAL.			
FT	CHAIN 20 1726 MEEROZOITE SURFACE PROTEIN 1.			
FT	CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).			
FT	CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).			

P47460 mycoplasma
P06101 saccharomyc
P16320 drosophila
P25805 plasmodium
P07228 gallus gall
O60879 homo sapien
Q99104 schizosacch
Q99104 mus musculu
P22620 plasmodium
O01969 thermotoga
P23745 plasmodium
P38822 saccharomyc

ALIGNMENTS

```

FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196197 MW; DD8AD45FA352BCF3 CRC64;

Query Match 100.0%; Score 2062; DB 1; Length 1726;
Best Local Similarity 100.0%; Pred. No. 9e-99;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGFENEVDVYKPLAGVYSLKKQIEKNFTFNLDNLNLSRLKRRKY 60
Db 1333 AISTVMDNLSGFENEVDVYKPLAGVYSLKKQIEKNFTFNLDNLNLSRLKRRKY 1392

Qy 61 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 1393 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 1452

Qy 121 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 180
Db 1453 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 1512

Qy 181 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 240
Db 1513 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 1572

Qy 241 LINDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 300
Db 1573 LINDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 1632

Qy 301 HLDRECKCLLNKQEGDKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTCTCK 360
Db 1633 HLDRECKCLLNKQEGDKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTCTCK 1692

Qy 361 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 394
Db 1693 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 1726

RESULT 2
MSPL_PLAFF
ID MSPL_PLAFF STANDARD; PRT; 1726 AA.
AC P50495;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (GP195).
GN MSP-1
OS Plasmodium falciparum (isolate Palo Alto / Uganda).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57270;
RN [1]
RX MEDLINE=89005525; PubMed=3049134;
RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
RA Stodolicki W.A.;
RT "Plasmodium falciparum: gene structure and hydrophathy profile of the
RT major merozoite surface antigen (gpi95) of the Uganda-Palo Alto
RT isolate."
RL Exp. Parasitol. 67:1-11(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -1- PPM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M37213; AAA29611.1;
DR InterPro; IPR000561;
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane, GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1726 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 827 827 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 924 924 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 990 990 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEFA2F9A026 CRC64;

Query Match 99.9%; Score 2059; DB 1; Length 1726;
Best Local Similarity 99.7%; Pred. No. 1.3e-98;
Matches 393; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGFENEVDVYKPLAGVYSLKKQIEKNFTFNLDNLNLSRLKRRKY 60
Db 1333 AISTVMDNLSGFENEVDVYKPLAGVYSLKKQIEKNFTFNLDNLNLSRLKRRKY 1392

Qy 61 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 1393 FLDVLESDLMQFKHISSEYIIESFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 1452

Qy 121 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 180
Db 1453 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 1512

Qy 181 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 240
Db 1513 YNNLVNKIDDDYLINLAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 1572

Qy 241 LINDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 300
Db 1573 LINDTKKMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 1632

Qy 301 HLDRECKCLLNKQEGDKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTCTCK 360
Db 1633 HLDRECKCLLNKQEGDKCVENPNTCNENNGCCADAKCTEEDSGSNGKKTCTCK 1692

Qy 361 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 394
Db 1693 PDSYPLFDGIFCSSSNFLGIFSLILMLILYSFI 1726

RESULT 3
MSPL_PLAFF
ID MSPL_PLAFF STANDARD; PRT; 1701 AA.
AC P13819;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA).
GN MSP-1.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

```

OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85142999; PubMed=2449612;
RA Peterson M.G., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens
of Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 27:291-302(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC
CC -1- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; M19143; AAA29653.1; -;
DR PIR; A54498; A54498.
DR InterPro; IPR000561; -;
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;

Query Match 98.38; Score 2027; DB 1; Length 1701;
Best Local Similarity 98.78; Pred. No. 5.9e-97;
Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGFENEYDVYILKPLAGVYRSLKKOIEKNFTNLDNLNLSRLKKRY 60
Db 1308 AISTVMDNLSGFENEYDVYILKPLAGVYRSLKKOIEKNFTNLDNLNLSRLKKRY 1367

Qy 61 FLDVLESLDMQFKHSSNEYIYEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 1368 FLDVLESLDMQFKHSSNEYIYEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 1427

Qy 121 YEKVLAKYKDDLSIKVIEKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 180
Db 1428 YEKVLAKYKDDLSIKVIEKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 1487

Qy 181 YNNLVNKIDDDYLLNKKAKINDCNVKEDEAHVKITKLSDLKAKIDDKIDLFKNHNDFEALKK 240
Db 1488 YNNLVNKIDDDYLLNKKAKINDCNVKEDEAHVKITKLSDLKAKIDDKIDLFKNHNDFEALKK 1547

Qy 241 LINDDTKKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKQKCPNSGCFR 300
Db 1548 LINDDTKKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKQKCPNSGCFR 1607

Qy 301 HLDRECKCLLNTYKQEGDKVCENPNPTCNENNGCDADATCTEEDSGSRKKTCTCK 360
Db 1548 HLDRECKCLLNTYKQEGDKVCENPNPTCNENNGCDADATCTEEDSGSRKKTCTCK 360

Db 1608 HLDRECKCLLNTYKQEGDKVCENPNPTCNENNGCDADATCTEEDSGSRKKTCTCK 1667

Qy 361 PDSYPLEDGFICSSSNFLGISFLILMLILYLSFI 394
Db 1668 PDSYPLEDGFICSSSNFLGISFLILMLILYLSFI 1701

RESULT 4
MSPL_PLAFM
ID MSPL_PLAFM STANDARD; PRT; 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite
Plasmodium falciparum."
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
Plasmodium falciparum merozoites: studies at the genetic level."
RL EMBO J. 4:3823-3829(1985).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC
CC -1- PFM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; X05624; CAA29112.1; -;
DR PIR; A26868; A26868.
DR PIR; B25120; B25120.
DR InterPro; IPR000561; -;
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).

```
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 98.3%; Score 2027; DB 1; Length 1701;
Best Local Similarity 98.7%; Pred. No. 5.5e-97;
Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AISVTMDNLSGFENEYDVYILKPLAGVYSLKKQIEKNFTFNLNLDILNSRLKRRKY 60
Db 1308 AISVTMDNLSGFENEYDVYILKPLAGVYSLKKQIEKNFTFNLNLDILNSRLKRRKY 1367
Qy 61 FLDVLESLDMQFKHISSEYIIEISFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 1368 FLDVLESLDMQFKHISSEYIIEISFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 1427
Qy 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPFLTNIETL 180
Db 1428 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPFLTNIETL 1487
Qy 181 YNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNHNDFAIKK 240
Db 1488 YNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNHNDFAIKK 1547
Qy 241 LINDTKMDLGLKLLSTGLVQNFPTTIISKLEGFQDMLNISQHCVKKOCPCNSGCFR 300
Db 1548 LINDTKMDLGLKLLSTGLVQNFPTTIISKLEGFQDMLNISQHCVKKOCPCNSGCFR 1607
Qy 301 HLDRECKCLLYNKQEGDKCVENPNTCNENNGCDADAKCTEEDSGSKKITCCTK 360
Db 1608 HLDRECKCLLYNKQEGDKCVENPNTCNENNGCDADAKCTEEDSGSKKITCCTK 1667
Qy 361 PDSVPLFDGIFCSSNPLGISFLILMLILYSFI 394
Db 1668 PDSVPLFDGIFCSSNPLGISFLILMLILYSFI 1701

RESULT 5
MSPI_PLAF3 STANDARD; PRT; 1682 AA.
AC P19598; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=8816657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen precursor p190 of Plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35727; AAA29715.1; -
DR EMBL; Y00087; CAA68280.1; -
DR EMBL; Z35326; CAA84555.1; -
DR PIR; S06286; S06286.
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82A1E159948CAD6 CRC64;

Query Match 98.2%; Score 2025.5; DB 1; Length 1682;
Best Local Similarity 99.0%; Pred. No. 6.5e-97;
Matches 390; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 1 AISVTMDNLSGFENEYDVYILKPLAGVYSLKKQIEKNFTFNLNLDILNSRLKRRKY 60
Db 1290 AISVTMDNLSGFENEYDVYILKPLAGVYSLKKQIEKNFTFNLNLDILNSRLKRRKY 1349
Qy 61 FLDVLESLDMQFKHISSEYIIEISFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 1350 FLDVLESLDMQFKHISSEYIIEISFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 1409
Qy 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPFLTNIETL 180
Db 1410 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPSPAKTDEQKESKFLPFLTNIETL 1468
Qy 181 YNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNHNDFAIKK 240
Db 1469 YNNLVNKIDYDYLINLAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNPYDFAIKK 1528
Qy 241 LINDTKMDLGLKLLSTGLVQNFPTTIISKLEGFQDMLNISQHCVKKOCPCNSGCFR 300
Db 1529 LINDTKMDLGLKLLSTGLVQNFPTTIISKLEGFQDMLNISQHCVKKOCPCNSGCFR 1588
Qy 301 HLDRECKCLLYNKQEGDKCVENPNTCNENNGCDADAKCTEEDSGSKKITCCTK 360
Db 1589 HLDRECKCLLYNKQEGDKCVENPNTCNENNGCDADAKCTEEDSGSKKITCCTK 1648
Qy 361 PDSVPLFDGIFCSSNPLGISFLILMLILYSFI 394
Db 1649 PDSVPLFDGIFCSSNPLGISFLILMLILYSFI 1682

RESULT 6
MSPI_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
```



```

FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 62.5%; Score 1289.5; DB 1; Length 1639;
Best Local Similarity 62.3%; Pred. No. 3.5e-59;
Matches 243; Conservative 54; Mismatches 72; Indels 21; Gaps 2;

QY 6 MDNLSGFENEYDVYIKPLAGVRSRKQIEKNFTFNLDNLNLSRLKRRKVFDDVL 65
DB 1270 IDNLSKIENEYVLYIKPLAGVRSRKQLENNWTFNVKDLNRSKRNKRFNVL 1329

QY 66 ESDLMQFKHSSNYYIEDSFKLNSQKNTLLSKYKIKESVENDIKFAQEGISYYEKV 125
DB 1330 ESDLPYKDLTSSNVVVKDPYKFLNKEKRDKEFLSSYVYKDSITDINFANDVLGYKIL 1389

QY 126 LAKYKDDLESIKKVIKEKEKFPSSPTTPSPAKTDEQKESKFLPFLNIETLYNNLV 185
DB 1390 SEKYSDLSISKKVI-----NDKQGENEKYLPFLNIETLYKTVN 1429

QY 186 NKIDYDLINKAKINDCNVEKDEAHVKITKLSDLKATDDIDLFKNHNDPEAIKKLINDD 245
DB 1430 DKIDLFVHLEAKVLYNTEKSNVEVKIKELNYLKTQDLADFKKNNFVGIADLSTDY 1489

QY 246 TKDMLGKLLSTGLV-QNFPNTIISKLEKGFQDMLNISQHCQVKKQCPNSGCFRHLDE 304
DB 1490 NHNNLTKFLSTGMVFENLAKTVLSNLDNLGMLNISQHCQVKKQCPNSGCFRHLDE 1549

QY 305 REECKCLLYKQEGDKCVENPPTCNENNGGCDADAKTEEDSGSKKTKCTCKPDSY 364
DB 1550 REECKCLLYKQEGDKCVENPPTCNENNGGCDADAKTEEDSGSKKTKCTCKPDSY 1609

QY 365 PLFDGIFCSSNFGISFLILMLILYSFI 394
DB 1610 PLFDGIFCSSNFGISFLILMLILYSFI 1639

RESULT 8
MSPL_PLAYO STANDARD; PRT; 1772 AA.
AC P13828;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (230 KDA).
GN MSP-1.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90205979; PubMed=2320061;
RA Lewis A.P.;
RT "Sequence analysis upstream of the gene encoding the precursor to the
major merozoite surface antigens of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 39:285-288(1990).
RN [2]
RP SEQUENCE OF 1093-1772 FROM N.A.
RC STRAIN=17XL;
RX MEDLINE=88124889; PubMed=2448778;
RA Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
RT "The 3' portion of the gene for a Plasmodium yoelii merozoite surface
antigen encodes the epitope recognized by a protective monoclonal
antibody.";
RT Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC -1- MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

```

```

CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03612; AAA29762.1; -
DR EMBL; J04668; AAA29702.1; -
DR PIR; A28121; A28121.
DR PIR; A45532; A45532.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT CHAIN 1 18 POTENTIAL.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 829 829 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1018 1018 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1090 1090 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1408 1408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1446 1446 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1541 1541 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1629 1629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1680 1680 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1521 1521 L -> V (IN REF. 2).
SQ SEQUENCE 1772 AA; 197230 MW; 9A6291658EB0F45D CRC64;

Query Match 29.8%; Score 613.5; DB 1; Length 1772;
Best Local Similarity 33.5%; Pred. No. 1.8e-24;
Matches 131; Conservative 90; Mismatches 143; Indels 27; Gaps 7;

QY 8 NILSGFENEYDVYIKPLAGVRSRKQIEKNFTFNLDNLNLSRLKRRKVFDDVL 67
DB 1401 DILSEFTNESLYVYTKRLGSTYKSLKXKMLREFSTIKEDMTNGLNKSQRNDFEVL 1460

QY 68 DLMQFKHSSNYYIEDSFKLNSQKNTLLSKYKIKESVENDIKFAQEGISYYEKVLA 127
DB 1461 ELDFKDLSTNKVIRNPYQLLNDKDKQIVNLVKATKGINEDIETTDGKIFKNMVE 1520

QY 128 KYKDDLESIKKVIKEKEKFPSSPTTPSPAKTDEQKE--SKFLPFLNIETLYNNLV 185
DB 1521 LYNQLAAVKEQIATIE-----AETNDTNKEKKKIPILEDLGLYETVI 1566

QY 186 NKIDYDLINKAKINDCNVEKDEAHVKITKLSDLKATDDIDLFKNHNDPEAIKKLINDD 245
DB 1567 GQAEVSEELQNRDLNRYKNEKAEFILTKNLEKYIOIDKLEFVEHAE-----NNKH 1619

QY 246 TKDMLGKLLSTGLV-QNFPNTIISKLEKGFQDMLNIS-QHOCV-KKQCPNSGCFRHL 302
DB 1620 IASTALNNLNKSGLVGEGESKILAKMLNMDGMDLLGVDPKHVCVDTDRIPKNAGCFROD 1679

QY 303 DERECKCLLYKQ-EGDKCVENPPTCNENNGGCDADAKTEEDSGSKKTKCTCKP 361
DB 1680 NGTEEWRCILGYKKRGEGTGVENNNPTCDINNGGCDPTASQNAESENSKTKICTCKP 1739

QY 362 DSYPLFDGIFCSSNFGISFLILMLILYS 392
DB 1740 TPNAVYEGVFCSSSFGSLIILITLIVEN 1770

RESULT 9
YL17_CAEEL
ID YL17_CAEEL STANDARD; PRT; 1130 AA.
AC Q11102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)

```


Qy 289 -KKQPCNSGCFRHLDERE-ECKCL-----LNYKQEGDKVE---NPNPTC 329
 Db 574 LKKQ-RAGTASFIDLTIEVETPLSLPSQDYILTSINAIDYEPEYKAMQVCGDSIIC 632
 Qy 330 NENGGCDADAKTEEDSGSGNGKKITCE 357
 Db 633 NTLN-----IAKDLKWKKGIRGKLVITTE 655

RESULT 11
 MYS2_DICDI
 ID MYS2_DICDI STANDARD; PRT: 2116 AA.
 AC P08799;
 DT 01-OCT-1988 (Rel. 09, Created)
 DT 01-NOV-1989 (Rel. 12, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Dictyostelium II HEAVY CHAIN, NON MUSCLE.
 GN MHCA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87092266; PubMed=3540939;
 RA Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
 RT "Conserved protein domains in a myosin heavy chain gene from
 Dictyostelium discoideum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
 RN [2]
 RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
 RC STRAIN=AX2;
 RX MEDLINE=90353583; PubMed=2387408;
 RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
 RA Gerisch G.;
 RT "Replacement of threonine residues by serine and alanine in a
 phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
 RL FEBS Lett. 269:239-243(1990).
 RN [3]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE=88112226; PubMed=2828113;
 RA Wagle G., Noegel A., Scheel J., Gerisch G.;
 RT "Phosphorylation of threonine residues on cloned fragments of the
 Dictyostelium myosin heavy chain.";
 RL FEBS Lett. 227:71-75(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345066; PubMed=7619795;
 RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
 RA Rayment I.;
 RT "X-ray structures of the myosin motor domain of Dictyostelium
 discoideum complexed with MgADP.BeF₃ and MgADP.ALF₄.";
 RL Biochemistry 34:8960-8972(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
 RX MEDLINE=95345067; PubMed=7619796;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
 truncated head of Dictyostelium discoideum myosin to 2.7-A
 resolution.";
 RL Biochemistry 34:8973-8981(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=96206189; PubMed=8611530;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
 Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
 RL Biochemistry 35:5404-5417(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
 RX MEDLINE=97452580; PubMed=9305951;
 RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
 RT "X-ray structures of the MgADP, MgATPgammaS, and MgAMPNP complexes
 of the Dictyostelium discoideum myosin motor domain.";

RL Biochemistry 36:11619-11628(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE=98070605; PubMed=9405148;
 RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
 RT "X-ray crystal structure and solution fluorescence characterization
 of Mg₂(3')-O-(N-methylanthraniloyl) nucleotides bound to the
 Dictyostelium discoideum myosin motor domain.";
 RL J. Mol. Biol. 274:394-407(1997).
 CC -|- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -|- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 CC (MLC-2).
 CC -|- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CC CORTEX.
 CC -|- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
 CC SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -|- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -|- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -|- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 CC POSITION (688).
 CC -|- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
 CC SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
 CC HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
 CC THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
 CC THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
 CC -|- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14628; AAA33227.1;
 DR PIR: A26655; A26655.
 DR PIR: S00250; S00250.
 DR PDB: 1MMA; 03-DEC-97.
 DR PDB: 1MKD; 17-AUG-96.
 DR PDB: 1MMG; 03-DEC-97.
 DR PDB: 1MMN; 03-DEC-97.
 DR PDB: 1MND; 17-AUG-96.
 DR PDB: 1MNE; 17-AUG-96.
 DR PDB: 1VOM; 23-DEC-96.
 DR PDB: 1LVK; 28-JAN-98.
 DR DictyDb; DD01008; mhca.
 DR InterPro; IPR000048;
 DR InterPro; IPR001609;
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00063; myosin_head; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KW Heptad repeat pattern; Methylation; Phosphorylation.
 FT DOMAIN 1 816 GLOBULAR HEAD (S1).
 FT DOMAIN 817 2116 RODLIKE TAIL (S2 AND LMM DOMAINS).
 FT NP_BIND 179 186 COILED COIL (POTENTIAL).
 FT DOMAIN 638 660 ACTIN-BINDING.
 FT DOMAIN 738 752 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
 FT MOD_RES 678 678 ALKYLATION (SH-1).
 FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
 FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).


```
FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770BB1EE56A1 CRC64;

Query Match 6.6%; Score 137; DB 1; Length 2116;
Best Local Similarity 21.9%; Pred. No. 6;
Matches 75; Conservative 63; Mismatches 132; Indels 72; Gaps 12;

QY 22 LKPLAGVYRSUKLOEKNIFFNLNLDILNSLRKSKRYFLDV-LESDLMOFKHISNEY 80
Db 1673 IKSILVAEVEDEKLEDEI-----LAKDLVKAKRALEVELEVRDGLDEEEDSR 1723
QY 81 IEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGISYEVKVLAKYKDDLESIRKVI 140
Db 1724 ELEDKRRLLTEVED-ISKYI---DAVEQNTKL-----DEAKKKLTDDVDTLKKQL 1771
QY 141 KEKEKEPPSPPTPPSPAKTDEQKSKFLPTNTIETLYNNLV-----NKIDDTLYLN 194
Db 1772 EDEKKKLNES-----ERAKKRLSENEFDLAKLDAEVKNRSRAEKDKRYEKDLKD 1822
QY 195 LKAKINDCNVEKDEAHVKITKLSDLKALDKIDLPKHNDFEATKKLINDTKKDLGKL 254
Db 1823 TRYKLNDEATKTTEIGAAL-----EODIELRSKLEQEQAKOAKDSKRTLEGEI 1876
QY 255 LSTGLVQNFNPNTIISKL-----IEGRFQOMLNISQHCYKQKCPENSGCFRHLDEREC 308
Db 1877 DNLRAQIEDCKIKRLEKEKRALEGELEL-----RETVEAEEDSKS-----EAEQS 1924
QY 309 KLLNYKQEGDKVCNPNPTCNENNGGCDADAKTEEDSGSN 350
Db 1925 KRLVELELEADAR-----RNLQKEIDAKEIAEDAKSN 1955

RESULT 12
RBPI_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBPI.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL Cell 69:1213-1226(1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M88097; AAA29743.1; -
CC HSP; P36956; IAM9.
CC Malaria; Receptor; Signal; Transmembrane.
CC SIGNAL 1 17 POTENTIAL.
CC CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
CC DOMAIN 18 2807 EXTRACELLULAR.
CC TRANSMEM 2808 2826 POTENTIAL.
CC DOMAIN 2827 2869 CYTOPLASMIC.
CC

FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;

Query Match 6.6%; Score 136.5; DB 1; Length 2869;
Best Local Similarity 21.7%; Pred. No. 8.8;
Matches 82; Conservative 70; Mismatches 145; Indels 81; Gaps 17;

QY 34 KOIEKNITFTNLNLDI-----LNSRLKKRYFLDVL-----ESDLMQFKHIS 76
Db 2265 RDIKKELYLFQHN-SDISIVEGGVQNNMLALYDKLNEEKREMDLYRISSETKLQMEHST 2323
QY 77 SNEYIIEEDSKLLSEQNTLLSKYIK-----ESVENDIKFAQEGI-----SY 121
Db 2324 DVFPMIELHKGMMETNKSLLKEKLLKSVNDHMHMEAMIKNGLUKTYPESVONTNNI 2383
QY 122 YEKVLAKYKDDLESIKK-----VKEKEKFPSPPTPPSPAKTDEQKSKFLPFL 174
Db 2384 YSVIEAEVK-TLEEIDRDYGDNYQIVVEHKKQFSILIDRTNALMDDIEIFPKENNYLME 2442
QY 175 TNETLNNLVNKIDDTLYLNKAKINDCNVEKDEAHVKITKLSDLKALDD-----KI 226
Db 2443 VNET-----IHRVNDYIEKITNKLVAQKTEYEQI-----LENIKQNDMDLQNIFLKKV 2491
QY 227 DLKHNHDFEAIKK-----LINDDTKDKMLGKLSTG-----LVONFPNTIISKLEGGKQ 277
Db 2492 SIIEY-----FENVKKKESILNDLYEQE-----RLKIGHELDIEIKRNVETLUSSYIDORME 2545
QY 278 DMLNISQHCYKQKCPENSGCFRHLDERECKCLLYNKQ--EGDKVCNPNPTCNENNGG 335
Db 2546 MW-----SKNLEKSKMMNYTSIYEL-EREANEINRAKQIKDDDTILNSVLEAAIQRGD 2601
QY 336 CDADAKTEEDSGSGNGK 353
Db 2602 MDAIFSQMSADRNPNEYK 2619

RESULT 13
YIO9_YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHEICAL 195.1 KDA PROTEIN IN DNA43-UBI1 INTERGENIC REGION.
GN Yli149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC 1994) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z38059; CAA86129.1; -
CC PIR; S48385; S48385.
CC SGI; S0001411; MLP2.
```

[illegible]

Query Match 6.5%; Score 135; DB 1; Length 1228;
Best Local Similarity 22.3%; Pred. No. 4.2;
Matches 75; Conservative 63; Mismatches 132; Indels 66; Gaps 15;

```
Y 29 YRSLK-KOIEKNIFTNLMNDILNRLKRYFLDVLSDLMQFKHSSNEYIEDSF 87
D 330 YQAMKLTULQKKNISUTVNDVRNT-----YSSLEGVSEDK-----SREF--QSLLK 376
Q 88 LNSQKNTLKSYYIKES---VENDIKFAQEGISYKVLAKYKODLESIKKVIKEEK 144
D 377 GLKSKSINVLRD--IVREQFKIQNDM--QETVAQLFKTVSSLSDELESTROIQKV 431
Q 145 EKFPSSPTTPPSAKTDEQKESKFL-----PFLNIETLYNNLVNKIDDIYLINKAK 198
D 432 ESVVSI-----AAQOKFVLQVQNRPTLTDLVLRNHNVRQEMTLT----- 473
Q 199 INDCNVEKDEAHVKITKLSDLKADDDKIDLPKNHD---FEAIKKLINDDTKDKMLGKL 254
D 474 ---CEKPKLELVQTHLEGA-----LEQHSRSILYIESLNKTLUS--KLKEVHEQL 520
Q 255 LSTGLVQNPNTIISKLEGFQDMLNISOHQVKKQCPENSGCFRHLDERECKCLINY 314
D 521 LSTEQVSDKNAPAAESVSNVTEYMS--TLHENIKKQSLMMLQMFEDLHIOESKINNTV 579
Q 315 KQEGDKVCNENPNTCENNGCGDADAKCTEDSGSN 350
D 580 SLEMEK--ESLRGECEDMLSKCRNDRFKQLKDTTEEN 613

RESULT 15
USOL_YEAST
ID USOL_YEAST STANDARD; PRT; 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE INTRACELLULAR PROTEIN TRANSPORT PROTEIN USOL.
GN USOL OR INTL OR IDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -!- SIMILARITY: BELONGS TO THE P115(TAP)/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; X54378; CA38253.1; -
DR EMBL; L03188; AAB00143.1; -
DR EMBL; U53668; AAB66659.1; -
DR PIR; A38455; A38455.
DR HSSP; P80220; IDIP.
DR SGD; S0002216; USOL.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil;
KW Calcium-binding.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CA_BIND 475 486 POTENTIAL.
FT CONFLICT 847 847 G -> E (IN REF. 2).
FT CONFLICT 924 924 E -> K (IN REF. 2).
FT CONFLICT 1253 1253 V -> I (IN REF. 2).
FT CONFLICT 1319 1319 I -> V (IN REF. 2).
FT CONFLICT 1461 1461 N -> S (IN REF. 2).
FT CONFLICT 1581 1581 G -> S (IN REF. 2).
FT CONFLICT 1600 1600 I -> V (IN REF. 2).
FT CONFLICT 1661 1661 R -> S (IN REF. 2).
FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 6.3%; Score 130.5; DB 1; Length 1790;
Best Local Similarity 24.3%; Pred. No. 11;
Matches 85; Conservative 50; Mismatches 106; Indels 109; Gaps 18;

QY 2 ISVTMDNILSGFENEYDVIYK--PLAGVYRSLKQKQIEKNFTNLMNDI--LNSRLKK 57
DB 1179 ISQLNDEITS--TQENESIKKKNDLEGEVKAAMKSTSEEQ---SNLKKSEIDALNLIQKE 1234

QY 58 RKYFLDVLESLDMO-----FKHISSENYIEDSFKLLNSEQKNT- 96
DB 1335 LKKKNEINSEASLLESIKSVESEVTKIKELQDCNFKKEYSE--LED--KLKASEDKNSK 1290

QY 97 ---LLKSYKIKESVE---NDIKFAQEGISYKVLAKYKDDLESIKKVIKEEK----- 144
DB 1291 YLELQKESKIKELDAKTTELKIQLEKITNLSKAKEKSELSRLAKTSSEERKNAEEQ 1350

QY 145 -EKPPSPPTTPPSPAKTDEQKESKFL-PFLNIETLYNNLVNKIDDIYLI-----NLK 196
DB 1351 LEKLNK-----IQKNQAFERKRLNNEGSTITQYSEKINTLEDELIRLQNEHELK 1404

QY 197 AKINDCNVEKDEAHVKIT-----KLSDLKADDDKIDLPK-----NHN 233
DB 1405 AKED-NTRSELEKVSLSNDELLEEKQNTIKSLQDEILSYKDKITRNDKILLSTERDNKR 1463

QY 234 DFEA-----IKKLINDPTK-----KMLGKLLST 257
DB 1464 DLESLEQLRAAQESKAKVEBGLKLEESSEKAELEKSKEMMKKLEST 1513
```

Search completed: August 8, 2001, 12:32:29
Job time: 197 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:30:49 ; Search time 43.63 seconds
(without alignments)
687.893 Million cell updates/sec

Title: US-09-500-376-3

Perfect score: 2052

Sequence: 1 AISTVMDNLSGFENEYDVI.....SNFLGSLFLLMLILYSFI 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2052	100.0	1701	2 A26868	major merozoite su
2	2033	99.1	1701	2 A54498	major merozoite su
3	2006	97.8	1726	1 SA20GM	major merozoite su
4	1987.5	96.9	651	2 S47282	merozoite surface
5	1987	96.8	1726	2 A45948	major merozoite su
6	1254.5	61.1	400	2 A45545	major merozoite su
7	1254.5	61.1	1639	2 S05603	major merozoite su
8	1133	55.2	1631	1 SA20K1	major merozoite su
9	1091	53.2	1640	2 A24594	probable major sur
10	826	40.3	1726	2 A39401	merozoite surface
11	815	39.7	1751	2 A45604	major blood-stage
12	593.5	28.9	680	2 A28121	major merozoite su
13	593.5	28.9	1772	2 A45532	major merozoite su
14	569	27.7	1785	2 A45546	major merozoite su
15	168.5	8.2	2401	2 T28676	rhoptry protein -
16	166	8.1	980	2 E71606	hypothetical prote
17	156.5	7.6	1939	2 T18372	repeat organellar
18	155.5	7.6	2166	2 G70163	hypothetical prote
19	153.5	7.5	1169	2 A64505	pl15 homolog - Met
20	153.5	7.5	1619	2 T18499	hypothetical prote
21	152	7.4	2269	2 T28677	rhoptry protein -
22	150	7.3	1187	2 T18355	hypothetical prote
23	149.5	7.3	1127	2 T28317	ORF MSV156 hypothe
24	147	7.2	1130	2 T34081	hypothetical prote
25	147	7.2	1679	2 S48385	hypothetical prote
26	146.5	7.1	1250	2 E81339	probable restricti
27	145.5	7.1	1712	2 C71618	hypothetical prote
28	145.5	7.1	2116	2 A26655	myosin heavy chain
29	142	6.9	1156	2 B70356	chromosome assembl

30 140 6.8 880 2 F75103 conserved hypothet
31 140 6.8 1005 2 A64465 hypothetical prote
32 139 6.8 1979 2 C71622 hypothetical prote
33 139 6.8 3394 2 T18501 hypothetical prote
34 138.5 6.7 671 2 H64502 hypothetical prote
35 138 6.7 793 2 E64545 surface membrane p
36 137 6.7 624 2 PC6003 hypothetical prote
37 137 6.7 821 2 S67087 hypothetical prote
38 136.5 6.7 819 2 E70105 pl15 protein homol
39 136 6.6 978 2 A70387 conserved hypothet
40 135.5 6.6 3119 2 T18414 protein g377 - mal
41 135 6.6 442 2 T18507 hypothetical prote
42 135 6.6 652 2 B59102 hypothetical prote
43 135 6.6 1079 2 T18356 membrane protein p
44 135 6.6 1313 2 F96673 hypothetical prote
45 135 6.6 1819 2 A71928 cag island protein

ALIGNMENTS

RESULT 1

A26868

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 09-Jun-2000

C:Accession: A26868

R:Fanabe, K.; Mackay, M.; Goman, M.; Scaife, J.G.

J. Mol. Biol. 195, 273-287, 1987

A:Title: Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmod
A:Reference number: A26868; MUID:88011243

A:Accession: A26868

A:Molecule type: DNA

A:Residues: 1-1701 <TAN>

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1701/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 100.0%; Score 2052; DB 2; Length 1701;
Best Local Similarity 100.0%; Pred. No. 9.1e-101;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AISTVMDNLSGFENEYDVIYLPAGVYSLKQIEKNLITNLDNLNLSLKKRY 60
DB 1308 AISTVMDNLSGFENEYDVIYLPAGVYSLKQIEKNLITNLDNLNLSLKKRY 1367
QY 61 FLDVLES DLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKIKESVENDIKFAQEGIS 120
DB 1368 FLDVLES DLMQFKHISSNEYIIEDSFKLLNSEQKNILLKSYKIKESVENDIKFAQEGIS 1427
QY 121 YEKVLAKYKDDLESIKVKIEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
DB 1428 YEKVLAKYKDDLESIKVKIEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNIETL 1487
QY 181 YNLVKNKIDYLLNLKAKINDCNVEXDEAHVTKLSDLKAIIDKIDLFKNTNDFEAIKK 240
DB 1488 YNLVKNKIDYLLNLKAKINDCNVEXDEAHVTKLSDLKAIIDKIDLFKNTNDFEAIKK 1547
QY 241 LINDYTKDMGLKLLSTGLVQIFPNTIISKLEGGKFDMLNISQHCVKKQCPENSCFR 300
DB 1548 LINDYTKDMGLKLLSTGLVQIFPNTIISKLEGGKFDMLNISQHCVKKQCPENSCFR 1607
QY 301 HLDERECKLLNKKOBDKCEENPNPTCNENGGCDADATCTEEDSGSSRKKITCETCK 360
DB 1608 HLDERECKLLNKKOBDKCEENPNPTCNENGGCDADATCTEEDSGSSRKKITCETCK 1667
QY 361 PDSYPLFDGIFCSSSNFLGLISFLLMLILYSFI 394
DB 1668 PDSYPLFDGIFCSSSNFLGLISFLLMLILYSFI 1701

```
RESULT 2
A:Residues: 1104-1726 <WEB2>
A:Cross-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmodi
A:Reference number: A54498; MUID:88142999
A:Accession: A54498
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PEW>
A:Cross-references: GB:M19143; NID:gl60412; PIDN:AAA29653.1; PID:gl60413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 99.1%; Score 2033; DB 2; Length 1701;
Best Local Similarity 99.2%; Pred. No. 9.1e-100;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AISVTMDNLSGFENEYDVIIYKPLAGVYRSKKQIEKNITFNLDNLNLSRLKRRKY 60
|||||
Db 1308 AISVTMDNLSGFENEYDVIIYKPLAGVYRSKKQIEKNITFNLDNLNLSRLKRRKY 1367
|||||

QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGIS 120
|||||
Db 1368 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGIS 1427
|||||

QY 121 YEKVLAKYKDDLESIKVKIKEKEKFPSPPTPPSPAKTDQKKESKFLPFLTNIELT 180
|||||
Db 1428 YEKVLAKYKDDLESIKVKIKEKEKFPSPPTPPSPAKTDQKKESKFLPFLTNIELT 1487
|||||

QY 181 YNNLVNKIDYLLNKAINDCNVKEDEAHVITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
|||||
Db 1488 YNNLVNKIDYLLNKAINDCNVKEDEAHVITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547
|||||

QY 241 LINDDTKMDLGLSTGLVQIFPNTIISKLEGGKQDMNLNISOHQCVKQKQCPENSGCFR 300
|||||
Db 1548 LINDDTKMDLGLSTGLVQIFPNTIISKLEGGKQDMNLNISOHQCVKQKQCPENSGCFR 1607
|||||

QY 301 HLDERECKLLNKKQEGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 360
|||||
Db 1608 HLDERECKLLNKKQEGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 1667
|||||

QY 361 PDSYPLFDGIFCSSNPLGISFLLILMLILYSFI 394
|||||
Db 1668 PDSYPLFDGIFCSSNPLGISFLLILMLILYSFI 1701
|||||

RESULT 3
SAZOGM
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
N:Alternate names: 195K glycoprotein
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 09-Jun-2000
C:Accession: A23386; S06361
R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.
Nucleic Acids Res. 14, 3311-3323, 1986
A:Title: Variation in the gene encoding a major merozoite surface antigen of the human m
A:Reference number: A23386; MUID:86205236
A:Accession: A23386
A:Molecule type: DNA
A:Residues: 1-1104 <WEB1>
A:Cross-references: EMBL:X03831
R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Wolff, R.
Nucleic Acids Res. 16, 1206, 1988
A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria pa
A:Reference number: S06361; MUID:88143999
A:Accession: S06361
A:Molecule type: DNA
```

```
A:Residues: 1104-1726 <WEB2>
A:Cross-references: EMBL:X03831
C:Comment: The merozoite stages of different strains have strain-specific surface ant
C:Species: Plasmodium falciparum
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jun-2000
C:Accession: A54498
R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
Mol. Biochem. Parasitol. 27, 291-302, 1988
A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmodi
A:Reference number: A54498; MUID:88142999
A:Accession: A54498
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1701 <PEW>
A:Cross-references: GB:M19143; NID:gl60412; PIDN:AAA29653.1; PID:gl60413
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 97.8%; Score 2006; DB 1; Length 1726;
Best Local Similarity 98.0%; Pred. No. 2.5e-98;
Matches 386; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 AISVTMDNLSGFENEYDVIIYKPLAGVYRSKKQIEKNITFNLDNLNLSRLKRRKY 60
|||||
Db 1333 AISVTMDNLSGFENEYDVIIYKPLAGVYRSKKQIEKNITFNLDNLNLSRLKRRKY 1392
|||||

QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGIS 120
|||||
Db 1393 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQKNTLLKSYKIKESVENDIKFAQEGIS 1452
|||||

QY 121 YEKVLAKYKDDLESIKVKIKEKEKFPSPPTPPSPAKTDQKKESKFLPFLTNIELT 180
|||||
Db 1453 YEKVLAKYKDDLESIKVKIKEKEKFPSPPTPPSPAKTDQKKESKFLPFLTNIELT 1512
|||||

QY 181 YNNLVNKIDYLLNKAINDCNVKEDEAHVITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
|||||
Db 1513 YNNLVNKIDYLLNKAINDCNVKEDEAHVITKLSDLKAIDDKIDLFKNTNDFEAIKK 1572
|||||

QY 241 LINDDTKMDLGLSTGLVQIFPNTIISKLEGGKQDMNLNISOHQCVKQKQCPENSGCFR 300
|||||
Db 1573 LINDDTKMDLGLSTGLVQIFPNTIISKLEGGKQDMNLNISOHQCVKQKQCPENSGCFR 1632
|||||

QY 301 HLDERECKLLNKKQEGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 360
|||||
Db 1633 HLDERECKLLNKKQEGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 1692
|||||

QY 361 PDSYPLFDGIFCSSNPLGISFLLILMLILYSFI 394
|||||
Db 1693 PDSYPLFDGIFCSSNPLGISFLLILMLILYSFI 1726
|||||

RESULT 4
S47282
merozoite surface antigen 1 - malaria parasite (Plasmodium falciparum) (strain RO-71)
C:Species: Plasmodium falciparum
A:Variety: strain RO-71
C:Date: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 09-Jun-2000
C:Accession: S47282
R:Tolle, R.; Bujard, H.; Cooper, J.A.
submitted to the EMBL Data Library, July 1994
A:Description: Plasmodium falciparum: recombination within the C-terminal region of m
A:Reference number: S47282
A:Accession: S47282
A:Molecule type: DNA
A:Residues: 1-651 <TOL>
A:Cross-references: EMBL:Z35329; NID:g535257; PIDN:CAA84558.1; PID:g535258
A:Experimental source: strain RO-71
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen

Query Match 96.9%; Score 1987.5; DB 2; Length 651;
Best Local Similarity 97.5%; Pred. No. 7.8e-98;
Matches 384; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 1 AISVTMDNLSGFENEYDVIIYKPLAGVYRSKKQIEKNITFNLDNLNLSRLKRRKY 60
|||||
Db 259 AISVTMDNLSGFENEYDVIIYKPLAGVYRSKKQIEKNITFNLDNLNLSRLKRRKY 318
|||||
```

QY 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGIS 120
|||||
Db 319 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGIS 378
|||||
QY 121 YEKVLAKYKDDLESIKKVIKEEKEPPSPPTTPPSPAKTDEQKESKFLPFTNIETL 180
|||||
Db 379 YEKVLAKYKDDLESIKKVIKEEKEPPSPPTTPPSPAKTDEQKESKFLPFTNIETL 437
|||||
QY 181 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKNTNDFEAIKK 240
|||||
Db 438 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKNTNDFEAIKK 497
|||||
QY 241 LINDDTKMDGLKLLSTGLVQIFPNTIISKLEKGFODMLNISQHCQVKKQCPNSGCFR 300
|||||
Db 498 LINDDTKMDGLKLLSTGLVQIFPNTIISKLEKGFODMLNISQHCQVKKQCPNSGCFR 557
|||||
QY 301 HLDRECKCLLNKQBGDKCEENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTK 360
|||||
Db 558 HLDRECKCLLNKQBGDKCEENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTK 617
|||||
QY 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
|||||
Db 618 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 651
|||||
RESULT 5
A45948
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
C:Species: Plasmodium falciparum
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jun-2000
C:Accession: A45948
R:Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
Exp. Parasitol. 67, 1-11, 1988
A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major mero
A:Reference number: A45948; MUID:89005525
A:Accession: A45948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1726 <CHA>
A:Cross-references: GB:M37213
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
Query Match 96.8%; Score 1987; DB 2; Length 1726;
Best Local Similarity 97.2%; Pred. No. 2.5e-97;
Matches 383; Conservative 1; Mismatches 10; Indels 0; Gaps 0;
QY 1 AISVTMDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLNLSRLKRRKY 60
|||||
Db 1333 AISVTMDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLNLSRLKRRKY 1392
|||||
QY 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGIS 120
|||||
Db 1393 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGIS 1452
|||||
QY 121 YEKVLAKYKDDLESIKKVIKEEKEPPSPPTTPPSPAKTDEQKESKFLPFTNIETL 180
|||||
Db 1453 YEKVLAKYKDDLESIKKVIKEEKEPPSPPTTPPSPAKTDEQKESKFLPFTNIETL 1512
|||||
QY 181 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKNTNDFEAIKK 240
|||||
Db 1513 YNNLVNKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKNTNDFEAIKK 1572
|||||
QY 241 LINDDTKMDGLKLLSTGLVQIFPNTIISKLEKGFODMLNISQHCQVKKQCPNSGCFR 300
|||||
Db 1573 LINDDTKMDGLKLLSTGLVQIFPNTIISKLEKGFODMLNISQHCQVKKQCPNSGCFR 1632
|||||
QY 301 HLDRECKCLLNKQBGDKCEENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTK 360
|||||
Db 1633 HLDRECKCLLNKQBGDKCEENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTK 1692
|||||

QY 361 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
|||||
Db 1693 PDSYPLFDGIFCSSSNFLGISFLLILMLILYSFI 1726
|||||
RESULT 6
A45545
major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments
C:Species: Plasmodium falciparum
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C:Accession: A45545
R:Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.
Mol. Biochem. Parasitol. 49, 29-33, 1991
A:Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protei
A:Reference number: A45545; MUID:92131048
A:Accession: A45545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-400 <BLA>
A:Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBIIP:77621)
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
Query Match 61.1%; Score 1254.5; DB 2; Length 400;
Best Local Similarity 61.2%; Pred. No. 2.1e-59;
Matches 240; Conservative 54; Mismatches 73; Indels 25; Gaps 3;
QY 6 MDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLNLSRLKRRKYFLDVL 65
|||||
Db 31 IDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLNLSRLKRRKYFLDVL 90
|||||
QY 66 ESDLMQFKHISSEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGISYVEKV 125
|||||
Db 91 ESDLMQFKHISSEYIIEDSFLLNSEQKNTLLKSKYIKESVENDIKFAQEGISYVEKV 150
|||||
QY 126 LAKYKDDLESIKKVIKEEKEPPSPPTTPPSPAKTDEQKESKFLPFTNIETLYKTNV 185
|||||
Db 151 SEKYSDDLSDIKKYI-----NDKQGENEKYLPFLNNIETLYKTNV 190
|||||
QY 186 NKIDYDYLNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLKNTNDFEAIKKLINDD 245
|||||
Db 191 DKIDLFVILHLEAKVNLVYKESVNEVKEIKELNYLKTIDQLADFKKNNNFVGIADLSTDY 250
|||||
QY 246 TKDMLGKLLSTGLVQIFPN---TIISKLEKGFODMLNISQHCQVKKQCPNSGCFRHL 302
|||||
Db 251 NNNLLTKFLSTGMV--FENLAKTVLSNLDGNLQGLMLNISQHCQVKKQCPNSGCFRHL 308
|||||
QY 303 DERECKCLLNKQBGDKCEENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTKPD 362
|||||
Db 309 DERECKCLLNKQBGDKCEENPNPTCNENNGGCDADATCTEEDSGSRKKITCECTKPD 368
|||||
QY 363 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 394
|||||
Db 369 SYPLFDGIFCSSSNFLGISFLLILMLILYSFI 400
|||||
RESULT 7
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
submitted to the EMBL Data Library, April 1989
A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:g9896; PIDN:CAA33163.1; PID:g9897
R:Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989


```
QY 186 NKIDYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF----KNTNDFEAIKKL 241
Db 475 GOAEYSEELQNRJDNKYKNAEFEILTKNLEKVIQIDKLDVEHAENKHTIASI--A 532

QY 242 INDNTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNIS-OHQCV-KKQCPNSGCF 299
Db 533 LNNLKSGLVGESES-----KKILAKMLNMDGMLLGVDPKHVCVDRDIPKNGCF 584

QY 300 RHLDERECKLLNYKO-EGDKCEENPNPTCNENNGGCDADATCTEEDSSSRKKITCEC 358
Db 585 RDDNGTEWRCLLGYKKGEGTGVENNNPTCDINNGGCDPTASQNAESTENSKKIIC 644

QY 359 TKPDSYPLFDGIFCSCSSNFIIGISFLILMLILYS 392
Db 645 KEPTPNAYIEGVFCSSSFMGLSILLITLIVFN 678

RESULT 13
A45532
major merozoite surface antigen precursor - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 26-Aug-1999
C:Accession: A45532; A45531
R:Lewis, A.P.
Mol. Biochem. Parasitol. 36, 271-282, 1989
A:Title: Cloning and analysis of the gene encoding the 230-kilodalton merozoite surface
A:Reference number: A45532; MUID:90014981
A:Accession: A45532
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1772 <LEW>
A:Cross-references: GB:J04668; NID:g160492; PID:g160493
R:Daly, T.M.; Burns Jr., J.M.; Long, C.A.
Mol. Biochem. Parasitol. 36, 283-285, 1989
A:Title: Precursor to the major merozoite surface antigen of Plasmodium yoelii: cloning
A:Reference number: A45531; MUID:90014982
A:Accession: A45531
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 454-1094 <DAL>
A:Cross-references: GB:J03975; NID:g160081; PID:g160082
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen

Query Match 28.98; Score 593.5; DB 2; Length 1772;
Best Local Similarity 32.58; Pred. No. 8e-24;
Matches 128; Conservative 96; Mismatches 137; Indels 33; Gaps 8;

QY 8 NILSGFENEYDVIYKPLAGVYRSKQIEKNIITFNLNLDILNLSLKRKYFLDVLES 67
Db 1401 DILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDTNGLNKSKQRNDFLEVL 1460

QY 68 DLMQFKHISSENYIEDSFKLLNSEQNILLKSYKIKESVENDIKFAQEGISYEVKLA 127
Db 1461 ELDLFKDLSTNKYVIRNPYQLLDNDKDKQIVNLKYATKGINEDIETTDGIKFFENW 1520

QY 128 KYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKKE--SKFLPLNLIETLYNNLV 185
Db 1521 LYNQLAAVKEQIATIE-----AETNDNKKEKKYIPILEDLKLGYETVI 1566

QY 186 NKIDYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF----KNTNDFEAIKKL 241
Db 1567 GOAEYSEELQNRJDNKYKNAEFEILTKNLEKVIQIDKLDVEHAENKHTIASI--A 1624

QY 242 INDNTKKMDLGLKLLSTGLVQIFPNTIISKLEGFQDMLNIS-OHQCV-KKQCPNSGCF 299
Db 1625 LNNLKSGLVGESES-----KKILAKMLNMDGMLLGVDPKHVCVDRDIPKNGCF 1676

QY 300 RHLDERECKLLNYKO-EGDKCEENPNPTCNENNGGCDADATCTEEDSSSRKKITCEC 358
Db 1677 RDDNGTEWRCLLGYKKGEGTGVENNNPTCDINNGGCDPTASQNAESTENSKKIIC 1736
```

```
QY 359 TKPDSYPLFDGIFCSCSSNFIIGISFLILMLILYS 392
Db 1737 KEPTPNAYIEGVFCSSSFMGLSILLITLIVFN 1770

RESULT 14
A45546
major merozoite surface antigen precursor - Plasmodium chabaudi chabaudi
C:Species: Plasmodium chabaudi chabaudi
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-Aug-1999
C:Accession: A45546
R:Deleersnijder, W.; Hendrix, D.; Bendahman, N.; Hanegreefs, J.; Brijs, L.; Hamers-Ca
Mol. Biochem. Parasitol. 43, 231-244, 1990
A:Title: Molecular cloning and sequence analysis of the gene encoding the major meroz
A:Reference number: A45546; MUID:91218805
A:Accession: A45546
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1785 <DEL>
A:Cross-references: GB:M34947; NID:g160597; PID:g160598
C:Superfamily: major merozoite surface antigen
C:Keywords: surface antigen
```

```
Query Match 27.7%; Score 569; DB 2; Length 1785;
Best Local Similarity 33.0%; Pred. No. 1.6e-22;
Matches 130; Conservative 92; Mismatches 138; Indels 34; Gaps 11;
```

```
QY 8 NILSGFENEYDVIYKPLAGVYRSKQIEKNIITFNLNLDI---LNSRLKRRKYFLDV 64
Db 1415 DILDAPKSENEYIVTKSLGNTYKSFKKMLKE---FSMIKEDIMTGLNYKLEKRNDFLDV 1471

QY 65 LESDLMQFKHISSENYIEDSFKLLNSEQNILLKSYKIKESVENDIKFAQEGISYEVK 124
Db 1472 LSVELALFKDINTNKVFYKPNPYQLLDNDKDKQINILKYAIKGVTEDETATDGIIEFNK 1531

QY 125 VLAKYKDDLESIKKVIKEEKFPSSPTTPSPAKTDEQKESKFLPLTNIETLYNNL 184
Db 1532 MIELYKPLQNAVNEQI-----AAIGTEP----TDAEKK--KYAPIEDLKLGYETI 1576

QY 185 VNKIDYLLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKKLND 244
Db 1577 LNGAEFESELLQHKLENYKIEKAGFDILMANLEYIRIDEKLEDF-----VESAEK--NK 1629

QY 245 DTKKMDLGLKLLSTGLV-QIFPNTIISKLEGFQDMLNI-SQHCVKKQCPNSGCFRHL 302
Db 1630 HIASIALNNLNKSLGVTEGESKKILAKMLNMDMLLIGISNHNVCISTSTPDNAGCCFRYD 1689

QY 303 DERECECKLLNYKO--EGDKCEENPNPTCNENNGGCDADATCTEEDS--GSSRRKKITCEC 358
Db 1690 DGTTEWRCLLGFKKDDDDGRCVADAPVCNNNGGCDKNADCREVENTDRDPSKKIVCTC 1749

QY 359 TKPDSYPLFDGIFCSCSSNFIIGISFLILMLILYS 392
Db 1750 KEPTPNAYIAGVFCSSSFMGLSILLITLIVFN 1783

RESULT 15
T28676
rhostry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676; A45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular ma
A:Reference number: T20507; MUID:97077455
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
```


THIS PAGE BLANK (0370)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:31:56 ; Search time 60.73 Seconds
(without alignments)
376.342 Million cell updates/sec

Title: US-09-500-376-5
Perfect score: 1982
Sequence: 1 AVTTSVIDNLSKIENEYEV.....SNFLGISFLILMLLYSFI 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*

```
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1942	98.0	1654	6 AAP50777	Sequence of the p1
2	1818.5	91.8	1639	19 AAW54145	p. falciparum synt
3	1726.5	87.1	355	20 AAY03372	Merozoite surface
4	1726.5	87.1	355	20 AAY05832	Merozoite surface
5	1726.5	87.1	361	20 AAY09373	Merozoite surface
6	1726.5	87.1	361	20 AAY05833	Merozoite surface
7	1720.5	86.8	376	20 AAY09374	Modified merozoite
8	1720.5	86.8	376	20 AAY05834	Modified merozoite
9	1282	64.7	262	22 AAB37610	Merozoite surface
10	553	27.9	116	18 AAW36103	pFMSPI(pl9)A prote
11	553	27.9	116	18 AAW22592	pFMSPI(pl9)A prote

12	472	23.8	96	22 AAB37608	Merozoite surface
13	472	23.8	108	22 AAB37609	Merozoite surface
14	456.5	23.0	127	18 AAW22593	pFMSPI(pl9)S prote
15	456.5	23.0	127	18 AAW36102	pFMSPI(pl9)S prote
16	302	15.2	54	14 AAR41356	MSP1EGF2A EGF2-lik
17	283	14.3	54	14 AAR41357	MSP1EGF2B EGF2-lik
18	243	12.3	93	22 AAB37611	Merozoite surface
19	243	12.3	106	14 AAR41358	P. yoelii combined
20	198	10.1	49	14 AAR41359	MSP1EGF1A EGF1-lik
21	198	10.1	49	14 AAR41355	MSP1EGF1B EGF1-lik
22	186	9.4	350	21 AAY70278	Recombinant vaccin
23	155.5	7.8	3973	21 AAB18253	Plasmodium falcipa
24	154	7.8	2010	21 AAB18218	Plasmodium falcipa
25	149	7.5	507	21 AAB18173	Plasmodium falcipa
26	149	7.5	722	21 AAB18291	Plasmodium falcipa
27	149	7.5	980	21 AAB18294	Plasmodium falcipa
28	149	7.5	1346	21 AAB18236	Plasmodium falcipa
29	145	7.3	1714	21 AAB18275	Plasmodium falcipa
30	142	7.2	1254	11 AAR07503	Merozoite apical-en
31	142	7.2	1254	18 AAW24575	Merozoite apical-en
32	141	7.1	2500	21 AAB18272	Plasmodium falcipa
33	140	7.1	1308	21 AAB18167	Plasmodium falcipa
34	138.5	7.0	2013	21 AAB18265	Plasmodium falcipa
35	137	6.9	1121	21 AAB18241	Plasmodium falcipa
36	135	6.8	1979	21 AAB18171	Plasmodium falcipa
37	134	6.8	1150	21 AAY97001	S. cerevisiae esse
38	133	6.7	1712	21 AAB18205	Plasmodium falcipa
39	132	6.7	558	21 AAB18273	Plasmodium falcipa
40	132	6.7	1817	21 AAB18301	Plasmodium falcipa
41	131.5	6.6	469	20 AAY20058	B. burgdorferi ant
42	131	6.6	2539	21 AAB18198	Plasmodium falcipa
43	130	6.6	504	21 AAB18228	Plasmodium falcipa
44	129.5	6.5	481	20 AAY20047	B. burgdorferi ant
45	129.5	6.5	497	20 AAY20046	B. burgdorferi ant

ALIGNMENTS

RESULT 1
AAP50777
ID AAP50777 standard; Protein; 1654 AA.
AC AAP50777;
DT 30-SEP-1991 (first entry)
XX Sequence of the p195 protein of Plasmodium falciparum.
DE Sequence of the p195 protein of Plasmodium falciparum.
XX Malaria vaccine; epitope; antigen; immunogen.
OS Plasmodium falciparum.
XX EPI54454-A.
PN 11-SEP-1985.
PD 21-FEB-1985;
PF 85EP-0301173.
XX 26-SEP-1984; 84GB-0024340.
PR 22-FEB-1984; 84GB-0004692.
PR 21-FEB-1985; 85GB-0004429.
XX (WELL) WELLCOME FOUNDATION LTD.
PA Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
PI WPI; 1985-224845/37.
DR N-PSDB; AAN50530.
XX Cloned DNA sequence encoding plasmodium falciparum protein -
PT useful for expressing the protein for use in vaccines against
PT malaria

XX PS Claim 6; Fig 1; 51pp; English.

XX CC The sequence encoding the p195 protein of Plasmodium falciparum

CC (AAW50530) and a peptide comprising at least one of its epitopes

CC (see AAP50777) are claimed. Also claimed is a vaccine for inducing

CC immunity to malaria comprising the novel peptide or P195 or a

CC peptide comprising at least one epitope when derived from the new

CC DNA sequence, together with a carrier.

XX CC

XX SQ Sequence 1654 AA;

Query Match 98.0%; Score 1942; DB 6; Length 1654;

Best Local Similarity 98.7%; Pred. No. 1.7e-130;

Matches 372; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AVTTSVIDNLSKIENEYEVLYLKPLAGYVRSLLKQLENNVMTFNVNVDILNSRFNKR 60

Db 1278 avttsvidnlskieneveyevlylkplagvyrslkqlennvmtfnvkvdkilnsrfnkre 1337

QY 61 NFKNVLESDLIPYKDLTSSNVVVKOPYKFLNKEKRDKFLSSNYIKDSITDINFANDVL 120

Db 1338 nfkvnlesdlipykdltsnvvvykopykflnkekrdkflssnyikdsitdinfandvl 1397

QY 121 GYKILSEKYSKDSLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKTDLFVHLEAKV 180

Db 1398 gyykilsekyksdlsikkyindkqgenekyplpflnnietlyktvndkldlfvhlleakv 1457

QY 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGM 240

Db 1458 lnyteksnvevkielnylktioqkladfknnnfvgiadlstdynhnnltkflstgm 1517

QY 241 VFENLLKSVLSNLDWKLARYVKHFTTPMRKKTMTIOQSSGCFRHLDERECKCLLYKOE 300

Db 1518 vfenllksvlsnldwklaryvkhfttpmrkktmtioqssgcfrrhldeereckcllnyke 1577

QY 301 GDKVNPNTCNENNGCDADAKTEEDSGNGKKITCECTKPCYPLDFGIFCSSSNF 360

Db 1578 gskvnsnptcnennngcdadakcteedsnggkkitcectkpcypmfvicsssnf 1637

QY 361 LGISFLLILMLILYSFI 377

Db 1638 lgisfllilmlilysfi 1654

RESULT 2

AAW54145

ID AAW54145 standard; Protein; 1639 AA.

XX AC AAW54145;

XX DT 23-SEP-1998 (first entry)

XX DE P. falciparum synthetic gp190 protein.

XX KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;

XX KW monoclonal antibody; passive immunisation; parasite.

XX OS Plasmodium falciparum.

XX OS Synthetic.

XX PN WO9814583-A2.

XX PD 09-APR-1998.

XX PF 02-OCT-1997; 97WO-EP05441.

XX PR 02-OCT-1996; 96DE-4040817.

XX PA (BUJA/) BUJARD H.

XX PI Bujard H, Pan W, Tolle R;

XX WPI: 1998-240088/21.

DR N-PSDB; AAV21451, AAV35363.

XX PT Recombinant production of complete gp190/MSP-1 Plasmodium surface

PT protein - useful in anti-malaria vaccines, also stabilising genes by

PT reducing their AT content

XX PS Example 1; Fig 3c; 48pp; German.

XX CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1

CC (merozoite surface) protein. The gene encoding this protein has been

CC stabilised by reducing the AT content of the nucleotide sequence. Such a

CC protein is useful in vaccines against malaria or for producing monoclonal

CC antibodies (for passive immunisation). The complete gp190 protein can now

CC be produced outside the parasite and has, at least over extended regions,

CC the native pattern of folding. Larger amounts of the protein can be

CC produced recombinantly than would be possible using the parasites as

CC source.

XX SQ Sequence 1639 AA;

Query Match 91.8%; Score 1818.5; DB 19; Length 1639;

Best Local Similarity 93.4%; Pred. No. 1.2e-121;

Matches 352; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEYEVLYLKPLAGYVRSLLKQLENNVMTFNVNVDILNSRFNKR 60

Db 1264 avttsvidnlskieneveyevlylkplagvyrslkqlennvmtfnvkvdkilnsrfnkre 1323

QY 61 NFKNVLESDLIPYKDLTSSNVVVKOPYKFLNKEKRDKFLSSNYIKDSITDINFANDVL 120

Db 1324 nfkvnlesdlipykdltsnvvvykopykflnkekrdkflssnyikdsitdinfandvl 1383

QY 121 GYKILSEKYSKDSLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKTDLFVHLEAKV 180

Db 1384 gyykilsekyksdlsikkyindkqgenekyplpflnnietlyktvndkldlfvhlleakv 1443

QY 181 LNYTEKSNVEVKIKELNYLKTIOQKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGM 240

Db 1444 lnyteksnvevkielnylktioqkladfknnnfvgiadlstdynhnnltkflstgm 1503

QY 241 VFENLLKSVLSNLDWKLARYVKHFTTPMRKKTMTIOQSSGCFRHLDERECKCLLYKOE 300

Db 1504 vfenllksvlsnldwklaryvkhfttpmrkktmtioqssgcfrrhldeereckcllnyke 1562

QY 301 GDKVNPNTCNENNGCDADAKTEEDSGNGKKITCECTKPCYPLDFGIFCSSSNF 360

Db 1563 gdkvnsnptcnennngcdadakcteedsnggkkitcectkpcysplfdgicsssnf 1622

QY 361 LGISFLLILMLILYSFI 377

Db 1623 lgisfllilmlilysfi 1639

RESULT 3

AAW09372

ID AAW09372 standard; Protein; 355 AA.

XX AC AAW09372;

XX DT 31-AUG-1999 (first entry)

XX DE Merozoite surface protein MSP-1-42.

XX KW MSP-1; merozoite surface protein; malaria; vaccine;

XX KW protein engineering; protein expression; codon usage;

XX KW transgenic animal.

XX OS Plasmodium falciparum.

XX PN WO9920774-A2.


```

Db 181 lnytyeksnvevkielnylktiqgkladfkknfnvfgiadlstdynhnnlltkflstgm 240
QY 241 VFENLLKSVLSNLLDWKLARYVYKHFTTPMRKKTMTIQSSGCFRHLDERECKCLINLYKQE 300
Db 241 vfenlaktvlsnllldgnl-qgmlnisqhcvkqcpqsgcfrhldereeckclinykqe 299
QY 301 GDKCVENPNTCNENNGCDADAKCTEEDSGSNGKKITCECTKPCYPLFDGIFCS 356
Db 300 gdkcvenpntcnenngcdadadakcteedsgsngkkitcectkpsypfldgifcs 355

RESULT 5
ID AAY09373 standard; Protein; 361 AA.
XX AAY09373;
AC AAY09373;
DT 31-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX Plasmodium falciparum.
XX WO9920774-A2.
PN 29-APR-1999.
PD 20-OCT-1998; 98WO-US22226.
PF 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX Chen LH, Meade H;
PI WPI; 1999-288313/24.
DR P-PSDB; AAX56009.
XX Modified malarial protein for use in anti-malarial vaccines
PT Example; Fig 2; 35pp; English.
PS This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. A nucleic acid (see AAX56008)
CC encoding MSP-1-42 has been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals by reducing the AT content and removing
CC mRNA instability motifs. The invention allows expression of
CC MSP-1-42 in the milk of transgenic animals, and also provides a DNA
CC vaccine comprising a vector containing the altered MSP-1-42 nucleic
CC acid.
XX Sequence 361 AA;

Query Match 87.1%; Score 1726.5; DB 20; Length 361;
Best Local Similarity 93.3%; Pred. No. 6.3e-116;
Matches 332; Conservative 6; Mismatches 17; Indels 1; Gaps 1;

QY 1 AVTTSVIDNLSKIENEYEVLYKPLAGVYSLKQLENNVTNPNVYKDLNLSRENKRE 60
Db 1 avtpsvidnlskieneyevlykplagvyrslkqlennvmtfnvkvdkllnsfrnkre 60
QY 61 NFKNVLESDLIPYKDLTSSNYYVVDPPYKFLNKRKDRKFLSSYNIKDSIDTIDNFANDVL 120

```

```

Db 61 nfkvnlesdlipykdltsnnyvvdpykflnkekrdkflssynyikdsidtdinfandvl 120
QY 121 GYYKILSEKYSDDLSDIKKYINDKOGENEKYLPLFNNTIETLYKTVNDKIDLFVHLEAKV 180
Db 121 gyykilsesksdlsdlsdkkyindkqgenekylpflnnietlyktvndkidlfvhlakv 180
QY 181 LNYTYEKSNNVEVKIELNYLKTIOQKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 240
Db 181 lnytyeksnvevkielnylktiqgkladfkknfnvfgiadlstdynhnnlltkflstgm 240
QY 241 VFENLLKSVLSNLLDWKLARYVYKHFTTPMRKKTMTIQSSGCFRHLDERECKCLINLYKQE 300
Db 241 vfenlaktvlsnllldgnl-qgmlnisqhcvkqcpqsgcfrhldereeckclinykqe 299
QY 301 GDKCVENPNTCNENNGCDADAKCTEEDSGSNGKKITCECTKPCYPLFDGIFCS 356
Db 300 gdkcvenpntcnenngcdadadakcteedsgsngkkitcectkpsypfldgifcs 355

RESULT 6
AAY05833
ID AAY05833 standard; Protein; 361 AA.
XX AAY05833;
AC AAY05833;
DT 02-AUG-1999 (first entry)
DE Merozoite surface protein MSP-1-42.
KW MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX Plasmodium falciparum.
XX WO9920766-A2.
PN 29-APR-1999.
PD 20-OCT-1998; 98WO-US22225.
PF 15-MAY-1998; 98US-0085649.
PR 20-OCT-1997; 97US-0062592.
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX Chen LH, Meade H;
PI WPI; 1999-302742/25.
DR N-PSDB; AAX25587.
XX New modified recombinant nucleic acid sequences useful for producing
PT malarial DNA vaccine
PS Disclosure; Fig 2; 43pp; English.
CC This present sequence comprises a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The C-terminal end of the sequence is
CC modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
CC AAX25593) encoding MSP-1-42 have been modified according to a method
CC of the invention in order to improve expression in mammalian cells
CC and in transgenic animals. The invention provides modified
CC recombinant nucleic acid sequences and methods for increasing the
CC mRNA levels and protein expression of proteins that are difficult
CC to express in cell culture systems, especially mammalian cell
CC culture systems or in transgenic animals. The preferred difficult
CC protein candidates for expression are those derived from lower
CC organisms such as parasites, bacteria and viruses that have DNA
CC coding sequences of high AT content or which have mRNA instability
CC motifs or rare codons relative to the recombinant expression system

```


CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.

XX SQ Sequence 361 AA;

Query Match 87.1%; Score 1726.5; DB 20; Length 361;
 Best Local Similarity 93.3%; Pred. No. 6.3e-116; Indels 1; Gaps 1;
 Matches 332; Conservative 6; Mismatches 17;

Qy 1 AVTTSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKR 60
 Db 1 avtpsvdnlslskieneyevlylplagvyrsllkqlennvmtfnvkvdlinsrfnkre 60
 Qy 61 NFKNVLESDDLIPYKDLTSSNVVVDPPYFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 120
 Db 61 nfkvnlesdlipykdltsnvvvdkpyflnkekrdkflssnyikdsidtdinfandvl 120
 Qy 121 GYKILSEKYSDDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 180
 Db 121 gyykilsekyssdlidsikkyindkqgenekylpflnnietlyktvndkidlfvhlleakv 180
 Qy 181 LNYTYEKSNEVVKIKELNYLKTIOQKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGM 240
 Db 181 lnytyeksnevkvikelnylktioqkladfknnnfvgiadlstdynhnnlltkflstgm 240
 Qy 241 VFENLLKSVLSNLLDWKLARYVHFHTTPMRKKTMTIQSSGCFRHLDERECKCLLNYKQE 300
 Db 241 vfenllksvlsnllldgnl-qgmlnlsqbcvkqcpqsgcfrhldeereckclllykqe 299
 Qy 301 GDKVCVENPTCNENNGGCDADAKCTEDSGSNGKKITCECTKPCYPLFDGIFCS 356
 Db 300 gdkvcvenpntcnennggcdadakcteedsngskitcectkpcsyplfdgifcs 355

RESULT 7

AA09374
 ID AA09374 standard; Protein; 376 AA.

XX AC AA09374;

XX DT 31-AUG-1999 (first entry)

XX DE Modified merozoite surface protein MSP-1-42.

XX KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.

XX OS Plasmodium falciparum.
 OS Synthetic.

XX FH Key Location/Qualifiers
 FT Peptide 1..15

FT Peptide /note= "beta-casein signal peptide"
 FT Peptide 371..376

FT FT /note= "6xHis tag"

FT FT Misc-difference 197

FT FT /note= "Asn in native MSP-1-42 (N181Q mutation)"

FT FT Misc-difference 278

FT FT /note= "Asn in native MSP-1-42 (N262Q mutation)"

PN WO9920774-A2.

XX PD 29-APR-1999.

XX PF 20-OCT-1998; 98WO-US22226.

XX PR 15-MAY-1998; 98US-0085649.

XX PR 20-OCT-1997; 97US-0062592.

XX PA (GEN2) GENZYME TRANSGENICS CORP.

XX CH Chen LH, Meade H;
 XX PI
 XX DR WPI; 1999-288313/24.
 XX DR P-PSDB; AAX56008.

XX FT Modified malarial protein for use in anti-malarial vaccines

XX PS Example; Fig 11; 35pp; English.

XX CC The present sequence represents a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The sequence has been modified to include
 CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
 CC tag. In addition, N181Q and N262Q mutations have been introduced to
 CC eliminate 2 N-glycosylation sites. These modifications allow the
 CC MSP-1-42 protein to be expressed in the milk of transgenic mice.

XX SQ Sequence 376 AA;

Query Match 86.8%; Score 1720.5; DB 20; Length 376;
 Best Local Similarity 93.0%; Pred. No. 1.8e-115;
 Matches 331; Conservative 5; Mismatches 19; Indels 1; Gaps 1;

Qy 1 AVTTSVIDNLSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVVKDILNSRFNKR 60
 Db 16 avtpsvdnlslskieneyevlylplagvyrsllkqlennvmtfnvkvdlinsrfnkre 75
 Qy 61 NFKNVLESDDLIPYKDLTSSNVVVDPPYFLNKEKRDKFLSSNYIKDSIDTDINFANDVL 120
 Db 76 nfkvnlesdlipykdltsnvvvdkpyflnkekrdkflssnyikdsidtdinfandvl 135
 Qy 121 GYKILSEKYSDDLDSIKKYINDKQGENEKYLPFLNNIETLYKTVDNDKIDLFVHLEAKV 180
 Db 136 gyykilsekyssdlidsikkyindkqgenekylpflnnietlyktvndkidlfvhlleakv 195
 Qy 181 LNYTYEKSNEVVKIKELNYLKTIOQKLADFKNNNFVGIADLSTDYNNHNLTKFLSTGM 240
 Db 196 lgytyeksnevkvikelnylktioqkladfknnnfvgiadlstdynhnnlltkflstgm 255
 Qy 241 VFENLLKSVLSNLLDWKLARYVHFHTTPMRKKTMTIQSSGCFRHLDERECKCLLNYKQE 300
 Db 256 vfenllksvlsnllldgnlqgmlq-issqbcvkqcpqsgcfrhldeereckclllykqe 314
 Qy 301 GDKVCVENPTCNENNGGCDADAKCTEDSGSNGKKITCECTKPCYPLFDGIFCS 356
 Db 315 gdkvcvenpntcnennggcdadakcteedsngskitcectkpcsyplfdgifcs 370

RESULT 8

AA05834
 ID AA05834 standard; Protein; 376 AA.

XX AC AA05834;

XX DT 02-AUG-1999 (first entry)

XX DE Modified merozoite surface protein MSP-1-42.

XX KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.

XX OS Plasmodium falciparum.
 OS Synthetic.

XX FH Key Location/Qualifiers
 FT Peptide 1..15

FT FT /note= "beta-casein signal peptide"
 FT Peptide 371..376

FT FT /note= "6xHis tag"

QY 241 VFENLKRSLVNLNDWKL 258
 ||||| I-||||||| I
 Db 241 vfenlaktvlsnldgnl 258

RESULT 10
 AAW36103
 ID AAW36103 standard; Protein; 116 AA.
 XX AC AAW36103;
 XX DT 25-MAR-1998 (first entry)
 XX DE PfMSPI(p19)A protein sequence.
 XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
 XX OS Plasmodium falciparum.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 XX FT Region 1..95
 FT /note= "amino acids derived from P. falciparum MSP1 p19
 FT fragment"
 FT Region 96..116
 FT /note= "glycosylphosphatidylinositol anchoring sequence"
 XX PN WO9730158-A2.
 XX PD 21-AUG-1997.
 XX PF 14-FEB-1997; 97WO-FR00290.
 XX PR 14-FEB-1996; 96FR-0001822.
 XX PA (INSP) INST PASTEUR.
 PA (UYN) UNIV NEW YORK STATE.
 XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX WPI; 1997-425033/39.
 DR N-PSDB; AAT94550.
 XX PT Recombinant protein containing the merozoite surface protein-1 p19
 PT fragment - useful in anti-malarial vaccines, diagnosis and protein
 PT purification
 XX PS Disclosure; Fig 1B; 85pp; French.

XX CC This is the amino acid sequence of a recombinant protein comprising
 CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
 CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
 CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
 CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX SQ Sequence 116 AA;

Query Match 27.9%; Score 553; DB 18; Length 116;
 Best Local Similarity 93.4%; Pred. No. 1.3e-32;
 Matches 99; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 272 KTM1QSSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADAKTEEDSG 331
 | :|||||
 Db 11 kkqpsngcgrhldereeckllnykqegdkcvnptcnenngcdadakteedsg 70-
 QY 332 SNGKKITCECTKPCYPLFDGIFCSSNFLGISFLILMLILYSFI 377
 |||||

Db 71 sngkkitcectkpsypplfdgfcssnflgisflilmlilysfi 116

RESULT 11
 AAW22592
 ID AAW22592 standard; Protein; 116 AA.
 XX AC AAW22592;
 XX DT 25-MAR-1998 (first entry)
 XX DE PfMSPI(p19)A protein sequence.
 XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
 KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
 XX OS Plasmodium falciparum.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 XX FT Region 1..95
 FT /note= "amino acids derived from P. falciparum MSP1 p19
 FT fragment"
 FT Region 96..116
 FT /note= "glycosylphosphatidylinositol anchoring sequence"
 XX PN WO9730159-A2.
 XX PD 21-AUG-1997.
 XX PF 14-FEB-1997; 97WO-FR00291.
 XX PR 14-FEB-1996; 96FR-0001821.
 XX PA (INSP) INST PASTEUR.
 PA (UYN) UNIV NEW YORK STATE.
 XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
 PI Roth C;
 XX WPI; 1997-425034/39.
 DR P-PSDB; AAW22592.

XX PT Recombinant protein containing Plasmodium merozoite surface
 PT protein-1 p42 fragment - useful in antimalarial vaccines, also new
 PT antibodies for diagnosis and protein purification
 XX PS Disclosure; Fig 1B; 85pp; French.

XX CC This is the amino acid sequence of a recombinant protein comprising
 CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
 CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
 CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
 CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
 CC The recombinant protein can be used for the production of anti-malarial
 CC vaccines, where the p19 fragment provides a high level of protective
 CC immunity since it includes epitopes not presented in the p42 fragment.
 XX SQ Sequence 116 AA;

Query Match 27.9%; Score 553; DB 18; Length 116;
 Best Local Similarity 93.4%; Pred. No. 1.3e-32;
 Matches 99; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 272 KTM1QSSGCFRHLDERECKLLNYKQEGDKCVENPNPTCNENNGCDADAKTEEDSG 331
 | :|||||
 Db 11 kkqpsngcgrhldereeckllnykqegdkcvnptcnenngcdadakteedsg 70
 QY 332 SNGKKITCECTKPCYPLFDGIFCSSNFLGISFLILMLILYSFI 377
 |||||
 Db 71 sngkkitcectkpsypplfdgfcssnflgisflilmlilysfi 116

```

RESULT 12
AAB37608
ID AAB37608 standard; protein; 96 AA.
XX
AC AAB37608;
XX
DT 27-FEB-2001 (first entry)
XX
DE Merozoite surface protein-1.
XX
KW Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS Plasmodium falciparum.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-GB01558.
XX
PR 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX
WPI; 2001-015762/02.
XX
DR Novel variants of the C-terminal fragment of Plasmodium merozoite
XX surface protein-1, useful as vaccines for treating or preventing
XX malaria.
XX
PS Example 2; Page 48; 126pp; English.
XX
CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The present sequence is the
CC wild-type MSP-1 protein. This sequence was used to generate the variants
CC of the present invention. The non-natural variants of the present
CC invention are useful for immunising a mammal against malaria, and can be
CC used to treat malaria.
XX
SQ Sequence 96 AA;

Query Match 23.88; Score 472; DB 22; Length 96;
Best Local Similarity 93.2%; Pred. No. 6.5e-27;
Matches 82; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 272 KTMIQSSGCFRHLDERECKCLLNKYQEGDKCVENPNPTCNENNGCDADAKTEEDSG 331
DB 9 kkgcpqnsqgrhldereeckcllnykgqdkcvnnpntcnennngcdadakteedsg 68

QY 332 SNGKKITCECTKPCYPLFDGIFCSSN 359
DB 69 sngkkitcectkpdspylfdgifcssn 96

RESULT 13
AAB37609
ID AAB37609 standard; Protein; 108 AA.
XX
AC AAB37609;
XX
DT 27-FEB-2001 (first entry)
XX

```

```

DE Merozoite surface protein-119.
XX
KW Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS Plasmodium falciparum.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-GB01558.
XX
PR 20-APR-1999; 99GB-0009072.
PR 13-MAY-1999; 99US-0311817.
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX
WPI; 2001-015762/02.
XX
DR N-PSDB; AAC68977.
XX
DR Novel variants of the C-terminal fragment of Plasmodium merozoite
XX surface protein-1, useful as vaccines for treating or preventing
XX malaria.
XX
PS Example 5; Fig 15; 126pp; English.
XX
CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.1.9. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is MSP-119 protein.
XX
SQ Sequence 108 AA;

Query Match 23.88; Score 472; DB 22; Length 108;
Best Local Similarity 93.2%; Pred. No. 7.6e-27;
Matches 82; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 272 KTMIQSSGCFRHLDERECKCLLNKYQEGDKCVENPNPTCNENNGCDADAKTEEDSG 331
DB 21 kkgcpqnsqgrhldereeckcllnykgqdkcvnnpntcnennngcdadakteedsg 80

QY 332 SNGKKITCECTKPCYPLFDGIFCSSN 359
DB 81 sngkkitcectkpdspylfdgifcssn 108

RESULT 14
AAB22593
ID AAB22593 standard; Protein; 127 AA.
XX
AC AAB22593;
XX
DT 25-MAR-1998 (first entry)
XX
DE PfMSP1(pl9)S protein sequence.
XX
KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; pl9;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
FH Key Location/Qualifiers
FT Peptide 1..19

```

```

FT Protein /note= "signal peptide"
FT 20...127 /note= "mature protein"
FT 1...32 /note= "derived from P. vivax MSP1"
FT Region /note= "encoded by restriction enzyme sequence used to
FT Region /note= "create the chimeric sequence"
FT 33...34 /note= "derived from P. vivax MSP1"
FT Region /note= "derived from P. falciparum C-terminal p19
FT 35...127 /note= "fragment of MSP1"
FT FT
FT FT
XX WO9730159-A2.
XX
XX 21-AUG-1997.
XX
XX 14-FEB-1997; 97WO-FR00291.
XX
XX 14-FEB-1996; 96FR-0001821.
XX
XX (INSP ) INST PASTEUR.
XX PA
XX (UYNV ) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX
XX WPI: 1997-425034/39.
XX P-PSDB; AAW22592.
XX
XX Recombinant protein containing Plasmodium merozoite surface
XX protein-1 p42 fragment - useful in antimalarial vaccines, also new
XX antibodies for diagnosis and protein purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the amino acid sequence of a chimeric protein comprising amino
XX acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
XX linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
XX C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX The recombinant protein can be used for the production of anti-malarial
XX vaccines, where the p19 fragment provides a high level of protective
XX immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 127 AA:
XX
XX Query Match 23.0%; Score 456.5; DB 18; Length 127;
XX Best Local Similarity 72.7%; Pred. NO. 1.2e-25;
XX Matches 80; Conservative 12; Mismatches 11; Indels 7; Gaps 1;
XX
XX QY 247 KSVLSNLLDWKLARYKHFTTPMRKKTMTIQSSGCFRHLDERECKCLLNKQEGDKCVE 306
XX Db 25 kqlvanvdefnshq-----gcvkkgcpensgcfhrldereeckcllnkqegdkcve 77
XX
XX QY 307 NPNPTCNENNGCDADAKCTEEDSGSNGKKITCECTKPCYPLFDGIFCS 356
XX Db 78 npnptcnennngcdadakteedsngskitcectkpcysyplfdgifcs 127
XX
XX RESULT 15
XX ID AAW36102
XX AC AAW36102 standard; Protein; 127 AA.
XX
XX AC AAW36102;
XX
XX 25-MAR-1998 (first entry)
XX
XX PfMSP1(p19)S protein sequence.
XX
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX

```

```

OS Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
XX Key Location/Qualifiers
XX Peptide 1...19 "signal peptide"
XX Protein 20...127
XX Region 1...32 "mature protein"
XX Region 33...34
XX Region /note= "derived from P. vivax MSP1"
XX Region /note= "encoded by restriction enzyme sequence used to
XX Region /note= "create the chimeric sequence"
XX Region 35...127
XX Region /note= "derived from P. falciparum C-terminal p19
XX fragment of MSP1"
XX
XX WO9730158-A2.
XX
XX 21-AUG-1997.
XX
XX 14-FEB-1997; 97WO-FR00290.
XX
XX 14-FEB-1996; 96FR-0001822.
XX
XX (INSP ) INST PASTEUR.
XX PA
XX (UYNV ) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX
XX WPI: 1997-425033/39.
XX N-PSDB; AAT94549.
XX
XX Recombinant protein containing the merozoite surface protein-1 p19
XX fragment - useful in anti-malarial vaccines, diagnosis and protein
XX purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the amino acid sequence of a chimeric protein comprising amino
XX acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
XX linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
XX C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX The recombinant protein can be used for the production of anti-malarial
XX vaccines, where the p19 fragment provides a high level of protective
XX immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 127 AA:
XX
XX Query Match 23.0%; Score 456.5; DB 18; Length 127;
XX Best Local Similarity 72.7%; Pred. NO. 1.2e-25;
XX Matches 80; Conservative 12; Mismatches 11; Indels 7; Gaps 1;
XX
XX QY 247 KSVLSNLLDWKLARYKHFTTPMRKKTMTIQSSGCFRHLDERECKCLLNKQEGDKCVE 306
XX Db 25 kqlvanvdefnshq-----gcvkkgcpensgcfhrldereeckcllnkqegdkcve 77
XX
XX QY 307 NPNPTCNENNGCDADAKCTEEDSGSNGKKITCECTKPCYPLFDGIFCS 356
XX Db 78 npnptcnennngcdadakteedsngskitcectkpcysyplfdgifcs 127
XX

```

Search completed: August 8, 2001, 12:31:57
Job time: 165 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:31:55 ; Search time 60.73 Seconds
(without alignments)
374.345 Million cell updates/sec

Title: US-09-500-376-4

Perfect score: 1967

Sequence: 1 AVTPSVIHNLKSIENEYEV.....SNFLGISFLLMLILYSFI 375

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*

```

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1901	96.6	1654	6 AAP50777	Sequence of the P1
2	1757.5	89.3	1639	19 AAW54145	P. falciparum synt
3	1665.5	84.7	355	20 AAY09372	Merozoite surface
4	1665.5	84.7	355	20 AAY05832	Merozoite surface
5	1665.5	84.7	361	20 AAY09373	Merozoite surface
6	1665.5	84.7	361	20 AAY05833	Merozoite surface
7	1659.5	84.4	376	20 AAY09374	Modified merozoite
8	1659.5	84.4	376	20 AAY05834	Modified merozoite
9	1246	63.3	262	22 AAB37610	Merozoite surface
10	528	26.8	116	18 AAW36103	PfMSP1(p19)A prote
11	528	26.8	116	18 AAW22592	PfMSP1(p19)A prote

12	447	22.7	96	22	AAB37608	Merozoite surface
13	447	22.7	108	22	AAB37609	Merozoite surface
14	432.5	22.0	127	18	AAW22593	PfMSP1(p19)S prote
15	432.5	22.0	127	18	AAW36102	PfMSP1(p19)S prote
16	278	14.1	54	14	AAW41356	MSPIEGF2A EGF2-lik
17	259	13.2	54	14	AAW41357	MSPIEGF2B EGF2-lik
18	233	11.8	93	22	AAB37611	Merozoite surface
19	232	11.8	106	14	AAW41358	P. yoelii combined
20	200	10.2	49	14	AAW41354	MSPIEGF1A EGF1-lik
21	197	10.0	49	14	AAW41355	MSPIEGF1B EGF1-lik
22	193	9.8	350	21	AAW70278	Recombinant vaccin
23	156	7.9	1714	21	AAB18275	Plasmodium falcipa
24	153	7.8	1346	21	AAB18236	Plasmodium falcipa
25	148	7.5	507	21	AAB18173	Plasmodium falcipa
26	146	7.4	722	21	AAB18291	Plasmodium falcipa
27	145	7.4	980	21	AAB18294	Plasmodium falcipa
28	142.5	7.2	2010	21	AAB18218	Plasmodium falcipa
29	141.5	7.2	3973	21	AAB18253	Plasmodium falcipa
30	139.5	7.1	1817	21	AAB18301	Plasmodium falcipa
31	139.5	7.1	2013	21	AAB18265	Plasmodium falcipa
32	137	7.0	504	21	AAB18228	Plasmodium falcipa
33	137	7.0	558	21	AAB18273	Plasmodium falcipa
34	136.5	6.9	1712	21	AAB18205	Plasmodium falcipa
35	136	6.9	2500	21	AAB18272	Plasmodium falcipa
36	135	6.9	1150	21	AAW97001	S. cerevisiae esse
37	132.5	6.7	1817	21	AAB18255	Plasmodium falcipa
38	132	6.7	1308	21	AAB18167	Plasmodium falcipa
39	131	6.7	469	20	AAW20058	B. burgdorferi ant
40	131	6.7	1182	21	AAB18288	Plasmodium falcipa
41	130.5	6.6	1121	21	AAB18241	Plasmodium falcipa
42	129.5	6.6	2206	21	AAB18254	Plasmodium falcipa
43	129.5	6.6	2485	21	AAB18172	Plasmodium falcipa
44	129	6.6	1254	11	AAW07503	Merozoite apical-en
45	129	6.6	1254	18	AAW24575	Merozoite apical-e

ALIGNMENTS

RESULT 1
AAP50777 standard; Protein: 1654 AA.
XX AC AAP50777;
XX DT 30-SEP-1991 (first entry)
XX DE Sequence of the P195 protein of Plasmodium falciparum.
XX KW Malaria vaccine; epitope; antigen; immunogen.
XX OS Plasmodium falciparum.
XX PN EP154454-A.
XX PD 11-SEP-1985.
XX PF 21-FEB-1985; 85EP-0301173.
XX PR 26-SEP-1984; 84GB-0024340.
XX PR 22-FEB-1984; 84GB-0004692.
XX PR 21-FEB-1985; 85GB-0004429.
XX (WELL) WELLCOME FOUNDATION LTD.
XX PA Holder A, Sandhu J, Odink K, Lockyer M, Riveros-Moreno V;
XX PI WPI; 1985-224845/37.
XX DR N-PSDB; AAN50530.
XX PT Cloned DNA sequence encoding plasmodium falciparum protein -
PT useful for expressing the protein for use in vaccines against
PT malaria

```
XX PS Claim 6; Fig 1; 5lpp; English.
XX CC The sequence encoding the P195 protein of Plasmodium falciparum
CC (AAN50530) and a peptide comprising at least one of its epitopes
CC (see AAP50777) are claimed. Also claimed is a vaccine for inducing
CC immunity to malaria comprising the novel peptide or P195 or a
CC peptide comprising at least one epitope when derived from the new
CC DNA sequence, together with a carrier.
XX SQ Sequence 1654 AA;

Query Match 96.6%; Score 1901; DB 6; Length 1654;
Best Local Similarity 97.1%; Pred. No. 1.6e-135;
Matches 366; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

Qy 1 AVTPSVTHNLSKTIENEYEVLYLKLPLAGVYRSLLKQLENNVMTFNVNKKDILNSPFNKRE 60
Db 1278 avttsvidnllsktieneyevlylklplagvyrsllkqlennvmtfnvnkdlnsrnfkre 1337

Qy 61 NFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKEKRDKFLSSYNYIKDSIDTDFNFANDVL 120
Db 1338 nfkvnlesdlipykdltsnyvvdpykflnkekrkflssynyikdsidtdinfandvl 1397

Qy 121 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKIDLFVHLEAKV 180
Db 1398 gyykilsekyksdlsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 1457

Qy 181 LNYTYEKSNEV--IKELIYLYKTIDQKLADFKNNFVGIADLSTDYNNHLLTKFLSTGM 238
Db 1458 lnytyeksnevki--ikeliylyktidqkladfknnfvgiadlstdynhlltkflstgm 1517:

Qy 239 VFENLLKSILSLNDWKLARYVKHFTTPMRKKTMIQNSGCFRHLDERECKCLLNKQK 298
Db 1518 vfenllksilslndwklaryvkhfttpmrkktmiqnsqcfrrhldeereckcllnkqk 1577

Qy 299 GSKCVENPNTCNENNGCDADAKCTEEDSGSNGKKITCQCTKPDSPYPLSMVIFCSSSNF 358
Db 1578 gskcvensnptcnennngcdadakteedsgsngkkitcctkpdpcplsmvifcsssnf 1637

Qy 359 LGISFLILMLILYSFI 375
Db 1638 lgisflilmlilysfi 1654

RESULT 2
AAN54145
ID AAN54145 standard; Protein; 1639 AA.
XX AC AAN54145;
XX XX 23-SEP-1998 (first entry)
XX DE
XX KW P. falciparum synthetic gp190 protein.
XX KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX KW monoclonal antibody; passive immunisation; parasite.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX PN WO9814583-A2.
XX PD 09-APR-1998.
XX XX
XX PF 02-OCT-1997; 97WO-EP05441.
XX PR 02-OCT-1996; 96DE-4040817.
XX PA (BUJA/) BUJARD H.
XX PI Bujard H, Pan W, Tolle R;
```

```
XX WPI; 1998-240088/21.
DR N-PSDB; AAV21451, AAV35363.
XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
PT protein - useful in anti-malaria vaccines, also stabilising genes by
PT reducing their AT content
XX Example 1; Fig 3c; 48pp; German.
PS
XX This sequence represents a modified Plasmodium falciparum gp190/MSP-1
CC (merozoite surface) protein. The gene encoding this protein has been
CC stabilised by reducing the AT content of the nucleotide sequence. Such a
CC protein is useful in vaccines against malaria or for producing monoclonal
CC antibodies (for passive immunisation). The complete gp190 protein can now
CC be produced outside the parasite and has, at least over extended regions,
CC the native pattern of folding. Larger amounts of the protein can be
CC produced recombinantly than would be possible using the parasites as
CC source.
XX SQ Sequence 1639 AA;

Query Match 89.3%; Score 1757.5; DB 19; Length 1639;
Best Local Similarity 91.2%; Pred. No. 1.2e-124;
Matches 344; Conservative 7; Mismatches 23; Indels 3; Gaps 2;

Qy 1 AVTPSVTHNLSKTIENEYEVLYLKLPLAGVYRSLLKQLENNVMTFNVNKKDILNSPFNKRE 60
Db 1264 avtpsvthnllsktieneyevlylklplagvyrsllkqlennvmtfnvnkdlnsrnfkre 1323

Qy 61 NFKNVLESDLIPYKDLTSSNVVVDKPYKFLNKEKRDKFLSSYNYIKDSIDTDFNFANDVL 120
Db 1324 nfkvnlesdlipykdltsnyvvdpykflnkekrkflssynyikdsidtdinfandvl 1383

Qy 121 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKIDLFVHLEAKV 180
Db 1384 gyykilsekyksdlsikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 1443

Qy 181 LNYTYEKSNEV--IKELIYLYKTIDQKLADFKNNFVGIADLSTDYNNHLLTKFLSTGM 238
Db 1444 lnytyeksnevki--ikeliylyktidqkladfknnfvgiadlstdynhlltkflstgm 1503

Qy 239 VFENLLKSILSLNDWKLARYVKHFTTPMRKKTMIQNSGCFRHLDERECKCLLNKQK 298
Db 1504 vfenllksilslndwklaryvkhfttpmrkktmiqnsqcfrrhldeereckcllnkqk 1562

Qy 299 GSKCVENPNTCNENNGCDADAKCTEEDSGSNGKKITCQCTKPDSPYPLSMVIFCSSSNF 358
Db 1563 gskcvensnptcnennngcdadakteedsgsngkkitcctkpdspylfdgfcsssnf 1622

Qy 359 LGISFLILMLILYSFI 375
Db 1623 lgisflilmlilysfi 1639

RESULT 3
AAY09372
ID AAY09372 standard; Protein; 355 AA.
XX AC AAY09372;
XX XX 31-AUG-1999 (first entry)
XX DE Merozoite surface protein MSP-1-42.
XX KW MSP-1; merozoite surface protein; malaria; vaccine;
XX KW protein engineering; protein expression; codon usage;
XX KW transgenic animal.
XX OS Plasmodium falciparum.
XX PN WO9920774-A2.
```



```
XX PD 29-APR-1999.
XX DE
XX PF 20-OCT-1998; 98WO-US222226.
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 97US-0062592.
XX PA (GENZ ) GENZYME TRANSGENICS CORP.
XX PI Chen LH, Meade H;
XX DR WPI; 1999-288313/24.
XX DR P-PSDB; AAX56008.
XX PT Modified malarial protein for use in anti-malarial vaccines
XX PS Example; Fig 1; 35pp; English.
XX CC The present sequence represents a 42 kDa C-terminal portion of
CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
CC important target for the development of a vaccine against
CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
CC has been modified (see AAX56008) compared to the native sequence (see
CC AAX56009) such that 306 nucleotide positions have been replaced to
CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
CC instability motifs while maintaining the same protein amino acid
CC sequence. These alterations allow MSP-1-42 to be expressed in
CC mammalian cell culture and in transgenic mice. Native MSP-1-12
CC is known to be difficult to express in cell culture systems,
CC mammalian cell culture systems or in transgenic animals. The
CC invention allows expression of MSP-1 protein in the milk of
CC transgenic animals, and also provides a DNA vaccine comprising a
CC vector containing the altered MSP-1-42 sequence.
XX SQ Sequence 355 AA;

Query Match 84.7%; Score 1665,5; DB 20; Length 355;
Best Local Similarity 91.0%; Pred. No. 1.5e-118;
Matches 324; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

QY 1 AVTPSVIHNLISKIENEYEVLYLKPLAGVYRSLLKQLENNVMTFNVVKDILNSPFNKR 60
DB 1 avtspvidnllskieneyevlylkplagvyrsllkqlennvmtfnvkvdklnsrfnkre 60
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKRDRFLSYNIKSIDTDINFANDVL 120
DB 61 nfkvnlesdlipykdltsnsvvvkdpkyfknkrdrkflssynyiksdtdinfandvl 120
QY 121 GYKILSEKYSKSDLSIKKYINDKQGENEYLPFLNNIETLYKTVDNKIDLVFVHLEAKV 180
DB 121 gyykilsekyksdlsikkyindkqgeneylpflnnietlyktvndkidl fvhleakv 180
QY 181 LNYTYEKSNEV--IKELIYLKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
DB 181 lnytyeksnevike lnylktiqliadfkknfnvgiadlst d ynnhlltkflstgm 240
QY 239 VFENLLSKILLSNLDLWKLYRVKFTTPMRKKTMIQONSGCFRHLDERECKCLLNKQE 298
DB 241 vfenlaktvisnllidgnl-qgmnlisqhcvkqcpqnsqcfrrhldeereckcllnykqe 299
QY 299 GSKCVENPPTCNENNGCDADAKCTEEDSGNSGKKTTCQCTKPDSPSLSVIFCS 354
DB 300 gdkcvenpntcnennngcdadacteedsgnsqgkktctcectkpdspylfdgifcs 355

RESULT 4
AAY05832
ID AAY05832 standard; Protein; 355 AA.
XX
AC AAY05832;
XX
```

```
DT 02-AUG-1999 (first entry)
XX
XX DE Merozoite surface protein MSP-1-42.
XX KW MSP-1; merozoite surface protein; malaria; vaccine;
XX KW protein engineering; protein expression; codon usage;
XX KW transgenic animal.
XX OS Plasmodium falciparum.
XX PN W09920766-A2.
XX PD 29-APR-1999.
XX PF 20-OCT-1998; 98WO-US222225.
XX PR 15-MAY-1998; 98US-0085649.
XX PR 20-OCT-1997; 97US-0062592.
XX PA (GENZ ) GENZYME TRANSGENICS CORP.
XX PI Chen LH, Meade H;
XX DR WPI; 1999-302742/25.
XX DR N-PSDB; AAX25586.
XX PT New modified recombinant nucleic acid sequences useful for producing
XX PT malarial DNA vaccine
XX PS Disclosure; Fig 1; 43pp; English.
XX CC The present sequence represents a 42 kDa C-terminal portion of
XX CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX CC important target for the development of a vaccine against
XX CC Plasmodium falciparum. The nucleic acid sequence encoding MSP-1-42
XX CC has been modified (see AAX25586) compared to the native sequence (see
XX CC AAX25587) such that 306 nucleotide positions have been replaced to
XX CC lower the AT content (from 76 to 49.7%) and to eliminate 10 mRNA
XX CC instability motifs while maintaining the same protein amino acid
XX CC sequence. These alterations allow MSP-1-42 to be expressed in
XX CC mammalian cell culture and in transgenic mice. The invention
XX CC provides modified recombinant nucleic acid sequences and methods for
XX CC increasing the mRNA levels and protein expression of proteins that
XX CC are difficult to express in cell culture systems, mammalian cell
XX CC culture systems or in transgenic animals. The preferred difficult
XX CC protein candidates for expression are those derived from lower
XX CC organisms such as parasites, bacteria and viruses that have DNA
XX CC coding sequences of high AT content or which have mRNA instability
XX CC motifs or rare codons relative to the recombinant expression system
XX CC to be used. The invention allows repression of MSP-1 protein in
XX CC the milk of transgenic animals, and also provides a DNA vaccine
XX CC comprising a vector containing the altered MSP-1-42 sequence.
XX SQ Sequence 355 AA;
```

```
Query Match 84.7%; Score 1665,5; DB 20; Length 355;
Best Local Similarity 91.0%; Pred. No. 1.5e-118;
Matches 324; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

QY 1 AVTPSVIHNLISKIENEYEVLYLKPLAGVYRSLLKQLENNVMTFNVVKDILNSPFNKR 60
DB 1 avtspvidnllskieneyevlylkplagvyrsllkqlennvmtfnvkvdklnsrfnkre 60
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKRDRFLSYNIKSIDTDINFANDVL 120
DB 61 nfkvnlesdlipykdltsnsvvvkdpkyfknkrdrkflssynyiksdtdinfandvl 120
QY 121 GYKILSEKYSKSDLSIKKYINDKQGENEYLPFLNNIETLYKTVDNKIDLVFVHLEAKV 180
DB 121 gyykilsekyksdlsikkyindkqgeneylpflnnietlyktvndkidl fvhleakv 180
QY 181 LNYTYEKSNEV--IKELIYLKTIQDKLADFKKNNFVGIADLSTDYNNHLLTKFLSTGM 238
```

```

||||| 181 lnyteksnvevkikelnlyktqkladfkknfnvgiadlstdynhnnlftstgm 240
||||| 239 VFENLLKSLNSLDDWKLARYVHFTTPMRKKTMIQNSGCFRHLDERECKCLLNKQE 298
||||| 241 vfenlaktvisnllldgnl-qgminisqhcvkkqcpnsgcfrhldereckccllnykqe 299
||||| 299 GSKCVENPNTCNENNGGCDADAKCTEEDSGSKKITCOCTKPDSPSLSMVIFCS 354
||||| 300 gdkvenpnptcnennngcdadacteedsgskkkitcctckpdsyplfdgifs 355

RESULT 5
AAV09373
ID AAV09373 standard; Protein; 361 AA.
AC AAV09373;
XX
XX 31-AUG-1999 (first entry)
XX
XX Merozoite surface protein MSP-1-42.
XX
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
XX Plasmodium falciparum.
XX
XX WO9920774-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98WO-US222226.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX
XX Chen LH, Meade H;
XX
XX WPI; 1999-288313/24.
XX P-PSDB; AAX56009.
XX
XX Modified malarial protein for use in anti-malarial vaccines
XX
XX Example; Fig 2; 35pp; English.
XX
XX This present sequence comprises a 42 kDa C-terminal portion of
XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The C-terminal end of the sequence is
XX modified to include a 6xHis tag. A nucleic acid (see AAX56008)
XX encoding MSP-1-42 has been modified according to a method
XX of the invention in order to improve expression in mammalian cells
XX and in transgenic animals by reducing the AT content and removing
XX mRNA instability motifs. The invention allows expression of
XX MSP-1-42 in the milk of transgenic animals, and also provides a DNA
XX vaccine comprising a vector containing the altered MSP-1-42 nucleic
XX acid.
XX
XX Sequence 361 AA;

Query Match 84.7%; Score 1665.5; DB 20; Length 361;
Best Local Similarity 91.0%; Pred. No. 1.5e-118;
Matches 324; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

QY 1 AVTPSVTHNTLSKTENEYEVLYLKPLAGVYRSLKQLKLENNVMTFNVNVKDLNSPKNRE 60
Db 1 avtpsvthntlskieniyevelykplagvyrsllkqlennvmtfnvkvdlnsrfrnk 60
QY 61 NFKNVLESDLIPYKDLTSSNVVVKDPYKFLNKEKROKFLSSYNIKDSIDTIDNFANDVL 120

```

```

||||| 61 nfkvnlesdlipykdltsnsvvkvdkyflnkekrdfllssynyikdsidtdinfandvl 120
||||| 121 GYVKILSEKYSKSLDSIKKYINDKQGENEKYLDPLNNIETLYKTVDNKIDLFVHLEAKV 180
||||| 121 gyykilsekyksldskkysikkyindkqgenekylpflnnietlyktvndkidlfvihleakv 180
||||| 181 LNTYKESNVE--IKELIYLYKTQDKLADFKKNNNFVGIADLSTDYNNHLLTKFLSTGM 238
||||| 181 lntyeksnvevkikelnlyktqkladfkknfnvgiadlstdynhnnlftstgm 240
||||| 239 VFENLLKSLNSLDDWKLARYVHFTTPMRKKTMIQNSGCFRHLDERECKCLLNKQE 298
||||| 241 vfenlaktvisnllldgnl-qgminisqhcvkkqcpnsgcfrhldereckccllnykqe 299
||||| 299 GSKCVENPNTCNENNGGCDADAKCTEEDSGSKKITCOCTKPDSPSLSMVIFCS 354
||||| 300 gdkvenpnptcnennngcdadacteedsgskkkitcctckpdsyplfdgifs 355

RESULT 6
AAV05833
ID AAV05833 standard; Protein; 361 AA.
XX
XX AAV05833;
XX
XX 02-AUG-1999 (first entry)
XX
XX Merozoite surface protein MSP-1-42.
XX
XX MSP-1; merozoite surface protein; malaria; vaccine;
KW protein engineering; protein expression; codon usage;
KW transgenic animal.
XX
XX Plasmodium falciparum.
XX
XX WO9920766-A2.
XX
XX 29-APR-1999.
XX
XX 20-OCT-1998; 98WO-US222225.
XX
XX 15-MAY-1998; 98US-0085649.
XX 20-OCT-1997; 97US-0062592.
XX
XX (GENZ ) GENZYME TRANSGENICS CORP.
XX
XX Chen LH, Meade H;
XX
XX WPI; 1999-302742/25.
XX N-PSDB; AAX25587.
XX
XX New modified recombinant nucleic acid sequences useful for producing
XX malarial DNA vaccine
XX
XX Disclosure; Fig 2; 43pp; English.
XX
XX This present sequence comprises a 42 kDa C-terminal portion of
XX malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
XX important target for the development of a vaccine against
XX Plasmodium falciparum. The C-terminal end of the sequence is
XX modified to include a 6xHis tag. Nucleic acids (see AAX25586 and
XX AAX25593) encoding MSP-1-42 have been modified according to a method
XX of the invention in order to improve expression in mammalian cells
XX and in transgenic animals. The invention provides modified
XX recombinant nucleic acid sequences and methods for increasing the
XX mRNA levels and protein expression of proteins that are difficult
XX to express in cell culture systems, especially mammalian cell
XX culture systems or in transgenic animals. The preferred difficult
XX protein candidates for expression are those derived from lower
XX organisms such as parasites, bacteria and viruses that have DNA
XX coding sequences of high AT content or which have mRNA instability
XX motifs or rare codons relative to the recombinant expression system

```

CC to be used. The invention allows expression of MSP-1 in the milk
 CC of transgenic animals, and also provides a DNA vaccine comprising a
 CC vector containing the altered MSP-1-42 nucleic acid.

XX SQ Sequence 361 AA;

Query Match 84.7%; Score 1665.5; DB 20; Length 361;
 Best Local Similarity 91.0%; Pred. No. 1.5e-118;
 Matches 324; Conservative 7; Mismatches 22; Indels 3; Gaps 2;

Qy 1 AVTPSVIHNLISKIENEYEVLYLKLPLAGVYRSKKQLNNVMTFNVNVDILNSPFPNKR 60
 Db 1 avtpsvihnliskieneyevlylklplagvyrslkqlennvmtfnvndilnsrfnkre 60
 Qy 61 NFKNVLESLLPYKDLSSNVVVDYKFLNKKRDKFLSSYNYIKDSIDTDFNFANDVL 120
 Db 61 nfkvnlesdlpykdlssnyvvdkpflnkkrdkflssnyikdsidtdinfandvl 120
 Qy 121 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKIDLFVHLEAKV 180
 Db 121 gyykilsekyskdsldskkyindkqgenekyplflnnietlyktvndkidlfvihleakv 180
 Qy 181 LNYTYEKSNE--IKELIYLTQDKLADFKKNNFVGIADLSTDYNNHNLITKFLSTGM 238
 Db 181 lnytyeksnevkeinelnylktiqdkladfkknfnvgiadlstdynhnnlltkflstgm 240
 Qy 239 VFENLLSKILSNLDMKLARYVHFTTPMRKKTMIQNSGCFRHLDERECKCLLNYKQE 298
 Db 241 vfenlaktvslnldgnlqgmnlshqcvkqcpnsgcfrhldereckcllnykqe 299
 Qy 299 GSKCVENPNTCNENNGCDADAKCTEEDSGSKKITCOCTKPDSPYPLSMVIFCS 354
 Db 300 gdkvenpnptcnennngcdadaktcedsgsngkkitcetckpdsyplfdgifics 355

RESULT 7

AY09374
 ID AAY09374 standard; Protein: 376 AA.

XX AC AAY09374;

XX DT 31-AUG-1999 (first entry)

XX DE Modified merozoite surface protein MSP-1-42.

KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.

XX OS Plasmodium falciparum.
 XX Synthetic.

XX FH Key Location/Qualifiers

FT Peptide 1..15
 FT /note= "beta-casein signal peptide"

FT Peptide 371..376

FT /note= "6xHis tag"

FT Misc-difference 197

FT /note= "Asn in native MSP-1-42 (N181Q mutation)"

FT Misc-difference 278
 /note= "Asn in native MSP-1-42 (N262Q mutation)"

XX FT WO9920774-A2.

XX PD 29-APR-1999.

XX XX 20-OCT-1998; 98WO-US22226.

XX XX 15-MAY-1998; 98US-0085649.

PR 20-OCT-1997; 97US-0062592.

XX XX (GENZ) GENZYME TRANSGENICS CORP.

XX PI Chen LH, Meade H;
 XX WI; 1999-288313/24.
 DR P-PSDB; AAX56008.
 XX Modified malarial protein for use in anti-malarial vaccines
 PT Example; Fig 11; 35pp; English.
 PS

XX The present sequence represents a 42 kDa C-terminal portion of
 CC malaria merozoite surface protein MSP-1 (i.e. MSP-1-42), an
 CC important target for the development of a vaccine against
 CC Plasmodium falciparum. The sequence has been modified to include
 CC an N-terminal beta-casein signal peptide and an N-terminal 6xHis
 CC tag. In addition, N181Q and N262Q mutations have been introduced to
 CC eliminate 2 N-glycosylation sites. These modifications allow the
 CC MSP-1-42 protein to be expressed in the milk of transgenic mice.

XX SQ Sequence 376 AA;

Query Match 84.4%; Score 1659.5; DB 20; Length 376;
 Best Local Similarity 90.7%; Pred. No. 4.5e-118;
 Matches 323; Conservative 6; Mismatches 24; Indels 3; Gaps 2;

Qy 1 AVTPSVIHNLISKIENEYEVLYLKLPLAGVYRSKKQLNNVMTFNVNVDILNSPFPNKR 60
 Db 16 avtpsvihnliskieneyevlylklplagvyrslkqlennvmtfnvndilnsrfnkre 75
 Qy 61 NFKNVLESLLPYKDLTSSNVVVDYKFLNKKRDKFLSSYNYIKDSIDTDFNFANDVL 120
 Db 76 nfkvnlesdlpykdltsnyvvdkpflnkkrdkflssnyikdsidtdinfandvl 135
 Qy 121 GYKILSEKYSKSDLSIKKYINDKQGENEKYLPFLNNIETLYKTVDNKIDLFVHLEAKV 180
 Db 136 gyykilsekyskdsldskkyindkqgenekyplflnnietlyktvndkidlfvihleakv 195
 Qy 181 LNYTYEKSNE--IKELIYLTQDKLADFKKNNFVGIADLSTDYNNHNLITKFLSTGM 238
 Db 196 lnytyeksnevkeinelnylktiqdkladfkknfnvgiadlstdynhnnlltkflstgm 255
 Qy 239 VFENLLSKILSNLDMKLARYVHFTTPMRKKTMIQNSGCFRHLDERECKCLLNYKQE 298
 Db 256 vfenlaktvslnldgnlqgmnlshqcvkqcpnsgcfrhldereckcllnykqe 314
 Qy 299 GSKCVENPNTCNENNGCDADAKCTEEDSGSKKITCOCTKPDSPYPLSMVIFCS 354
 Db 315 gdkvenpnptcnennngcdadaktcedsgsngkkitcetckpdsyplfdgifics 370

RESULT 8

AY05834
 ID AAY05834 standard; Protein: 376 AA.

XX AC AAY05834;

XX DT 02-AUG-1999 (first entry)

XX DE Modified merozoite surface protein MSP-1-42.

KW MSP-1; merozoite surface protein; malaria; vaccine;
 KW protein engineering; protein expression; codon usage;
 KW transgenic animal; mutant.

XX OS Plasmodium falciparum.
 XX Synthetic.

XX FH Key Location/Qualifiers

FT Peptide 1..15
 FT /note= "beta-casein signal peptide"

FT Peptide 371..376

FT /note= "6xHis tag"


```

Qy 239 VFENLKSILSNLLDWKL 256
    ||||| |::||| |
Db 241 vfenlaktvlsnllgdnl 258

RESULT 10
ID AAW36103 standard; Protein; 116 AA.
XX AC AAW36103;
XX DT 25-MAR-1998 (first entry)
XX DE PfMSPI(p19)A protein sequence.
XX KW Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Region 1..95
XX FT /note= "amino acids derived from P. falciparum MSP1 p19
XX FT fragment"
XX FT Region 96..116
XX FT /note= "glycosylphosphatidylinositol anchoring sequence"
XX PN WO9730158-A2.
XX PD 21-AUG-1997.
XX PF 14-FEB-1997; 97WO-FR00290.
XX PR 14-FEB-1996; 96FR-0001822.
XX PA (INSP ) INST PASTEUR.
XX PA (UYNV ) UNIV NEW YORK STATE.
XX PI Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX PI Roth C;
XX DR WPI; 1997-425033/39.
XX DR N-PSDB; AAT94550.
XX PT Recombinant protein containing the merozoite surface protein-1 p19
XX PT fragment - useful in anti-malarial vaccines, diagnosis and protein
XX PT purification
XX PS Disclosure; Fig 1B; 85pp; French.
XX CC This is the amino acid sequence of a recombinant protein comprising
XX CC amino acids 1613-1705 of the Plasmodium falciparum merozoite surface
XX CC protein 1 (MSP1) 19 kD C-terminal fragment (p19), linked to a
XX CC glycosylphosphatidylinositol membrane anchoring sequence. p19 is the
XX CC C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX CC The recombinant protein can be used for the production of anti-malarial
XX CC vaccines, where the p19 fragment provides a high level of protective
XX CC immunity since it includes epitopes not presented in the p42 fragment.
XX SQ Sequence 116 AA;

Query Match 26.8%; Score 528; DB 18; Length 116;
Best Local Similarity 90.6%; Pred. No. 6.le-33;
Matches 96; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 270 KTM1QQNSGCFRHLDERECKLLNYKQEGSKVCVENPNPTCNENNGCDADAKTEEDSG 329
    | :||||| |
Db 11 kkgqpsngcfrhldereeckllnykqegdkvcvenpnptcnenngcdadakteedsg 70

Qy 330 SNGKKITCQCTKPDSPYPLSMVIFCSSNFIIGISFLILMLILYSFI 375
    ||||| |
Db 71 sngkkitcectkpdspylfdgfcssnflglsflllmlilysfi 116

```

```

RESULT 12
AAB37608
ID AAB37608 standard; protein; 96 AA.
XX
AC AAB37608;
XX
DT 27-FEB-2001 (first entry)
XX
DE Merozoite surface protein-1.
XX
KW Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS Plasmodium falciparum.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-GB01558.
XX
PR 20-APR-1999; 99GB-0009072.
XX
PR 13-MAY-1999; 99US-0311817.
XX
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX
WPI; 2001-015762/02.
XX
DR N-PSDB; AAC68977.
XX
XX
PT Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria.
XX
XX
PS Example 2; Page 48; 126pp; English.
XX
XX
CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.9. The present sequence is the
CC wild-type MSP-1 protein. This sequence was used to generate the variants
CC of the present invention. The non-natural variants of the present
CC invention are useful for immunising a mammal against malaria, and can be
CC used to treat malaria.
XX
SQ Sequence 96 AA;

Query Match 22.7%; Score 447; DB 22; Length 96;
Best Local Similarity 89.8%; Pred. No. 6.6e-27;
Matches 79; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 270 KTM1QNSGCFRHLDERECKLLNYKQEGSKCVENPNTCNENNGCDADAKTEEDSG 329
DB 9 kkgcpqnsqgcfhrldereckllnykqegdkcvenpntcnennngcdadakteedsg 68

QY 330 SNGKKITCQCTKPDSPVLSWIFCSSN 357
DB 69 sngkkitcectkpdspvplfdgfcssn 96

RESULT 13
AAB37609
ID AAB37609 standard; Protein; 108 AA.
XX
AC AAB37609;
XX
DT 27-FEB-2001 (first entry)
XX
DE Merozoite surface protein-119.
XX
KW Merozoite surface protein; protazoacide; vaccine; malaria.
XX
OS Plasmodium falciparum.
XX
PN WO200063245-A2.
XX
PD 26-OCT-2000.
XX
PF 20-APR-2000; 2000WO-GB01558.
XX
PR 20-APR-1999; 99GB-0009072.
XX
PR 13-MAY-1999; 99US-0311817.
XX
PR 25-MAY-1999; 99CA-2271451.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaipibull C;
XX
WPI; 2001-015762/02.
XX
DR N-PSDB; AAC68977.
XX
XX
PT Novel variants of the C-terminal fragment of Plasmodium merozoite
PT surface protein-1, useful as vaccines for treating or preventing
PT malaria.
XX
XX
PS Example 5; Fig 15; 126pp; English.
XX
XX
CC The present invention relates to non-natural variants of a C-terminal
CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
CC non-natural variants have reduced affinity for at least 1 antibody
CC capable of blocking a second antibody that inhibits the proteolytic
CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
CC MSP-1.4.2, compared to natural MSP-1.9. The non-natural variants of the
CC present invention are useful for immunising a mammal against malaria, and
CC can be used to treat malaria. The present sequence is MSP-119 protein.
XX
SQ Sequence 108 AA;

Query Match 22.7%; Score 447; DB 22; Length 108;
Best Local Similarity 89.8%; Pred. No. 7.7e-27;
Matches 79; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 270 KTM1QNSGCFRHLDERECKLLNYKQEGSKCVENPNTCNENNGCDADAKTEEDSG 329
DB 21 kkgcpqnsqgcfhrldereckllnykqegdkcvenpntcnennngcdadakteedsg 80

QY 330 SNGKKITCQCTKPDSPVLSWIFCSSN 357
DB 81 sngkkitcectkpdspvplfdgfcssn 108

RESULT 14
AAB22593
ID AAB22593 standard; Protein; 127 AA.
XX
AC AAB22593;
XX
DT 25-MAR-1998 (first entry)
XX
DE PfMSP1(p19)S protein sequence.
XX
KW Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
KW Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX
OS Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
PH Key Location/Qualifiers
FT Peptide 1..19

```

```

FT      Protein      /note= "signal peptide"
FT      20..127
FT      /note= "mature protein"
FT      1..32
FT      Region      /note= "derived from P. vivax MSP1"
FT      33..34
FT      Region      /note= "encoded by restriction enzyme sequence used to
FT      35..127      create the chimeric sequence"
FT      /note= "derived from P. falciparum C-terminal p19
FT      fragment of MSP1"
XX
XX      WO9730159-A2.
XX
XX      21-AUG-1997.
XX
XX      14-FEB-1997; 97WO-FR00291.
XX
XX      14-FEB-1996; 96FR-0001821.
XX
XX      (INSP ) INST PASTEUR.
XX      (UYNV ) UNIV NEW YORK STATE.
XX
XX      Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX      Roth C;
XX
XX      WPI; 1997-425034/39.
XX      P-PSDB; AAW22592.
XX
XX      Recombinant protein containing Plasmodium merozoite surface
XX      protein-1 p42 fragment - useful in antimalarial vaccines, also new
XX      antibodies for diagnosis and protein purification
XX
XX      Disclosure; Fig 1C; 85pp; French.
XX
XX      This is the amino acid sequence of a chimeric protein comprising amino
XX      acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
XX      linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
XX      C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX      p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX      The recombinant protein can be used for the production of anti-malarial
XX      vaccines, where the p19 fragment provides a high level of protective
XX      immunity since it includes epitopes not presented in the p42 fragment.
XX
XX      Sequence 127 AA;

Query Match 22.08; Score 432.5; DB 18; Length 127;
Best Local Similarity 70.0%; Pred. No. 1.2e-25;
Matches 77; Conservative 12; Mismatches 14; Indels 7; Gaps 1;

QY 245 KSILSLDLWKLYRVKHTTPMKKTMIOQNSGCFRHLDERECKLLNYKQSGSKVE 304
Db 25 kqlvanvdefnlsqh-----qcvkkqcpensgcfhldereckcllnykqegdkcve 77

QY 305 NPNTCNENGGCDADAKCTEEDSGSGKKTCTCQTKPDSPVPLSMWIFCS 354
Db 78 npntcnennnggcdadakcteedsngskktctcctkpdspvplfdgifs 127

RESULT 15
AAW36102
ID AAW36102 standard; Protein; 127 AA.
XX
XX AAW36102;
XX
XX 25-MAR-1998 (first entry)
XX
XX PfMSP1(p19)S protein sequence.
XX
XX Chimeric; Plasmodium vivax; merozoite surface protein; MSP1; p19;
XX Plasmodium falciparum; malaria; vaccine; immunity; epitope.
XX

```

```

OS Chimeric - Plasmodium vivax.
OS Chimeric - Plasmodium falciparum.
XX
XX Key Location/Qualifiers
XX Peptide 1..19
XX Protein /note= "signal peptide"
XX 20..127
XX /note= "mature protein"
XX Region 1..32
XX /note= "derived from P. vivax MSP1"
XX Region 33..34
XX /note= "encoded by restriction enzyme sequence used to
XX 35..127      create the chimeric sequence"
XX /note= "derived from P. falciparum C-terminal p19
XX fragment of MSP1"
XX
XX WO9730158-A2.
XX
XX 21-AUG-1997.
XX
XX 14-FEB-1997; 97WO-FR00290.
XX
XX 14-FEB-1996; 96FR-0001822.
XX
XX (INSP ) INST PASTEUR.
XX (UYNV ) UNIV NEW YORK STATE.
XX
XX Barnwell JW, Longacre-Andre S, Mendis K, Nato F;
XX Roth C;
XX
XX WPI; 1997-425033/39.
XX N-PSDB; AAT94549.
XX
XX Recombinant protein containing the merozoite surface protein-1 p19
XX fragment - useful in anti-malarial vaccines, diagnosis and protein
XX purification
XX
XX Disclosure; Fig 1C; 85pp; French.
XX
XX This is the amino acid sequence of a chimeric protein comprising amino
XX acids 1-32 of the Plasmodium vivax merozoite surface protein 1 (MSP1)
XX linked to amino acids Asn1613-Ser1705 which correspond to the 19 kD
XX C-terminal fragment (p19) of MSP-1 from Plasmodium falciparum.
XX p19 is the C-terminal fragment of the 42 kD MSP1 from Plasmodium species.
XX The recombinant protein can be used for the production of anti-malarial
XX vaccines, where the p19 fragment provides a high level of protective
XX immunity since it includes epitopes not presented in the p42 fragment.
XX
XX Sequence 127 AA;

Query Match 22.08; Score 432.5; DB 18; Length 127;
Best Local Similarity 70.0%; Pred. No. 1.2e-25;
Matches 77; Conservative 12; Mismatches 14; Indels 7; Gaps 1;

QY 245 KSILSLDLWKLYRVKHTTPMKKTMIOQNSGCFRHLDERECKLLNYKQSGSKVE 304
Db 25 kqlvanvdefnlsqh-----qcvkkqcpensgcfhldereckcllnykqegdkcve 77

QY 305 NPNTCNENGGCDADAKCTEEDSGSGKKTCTCQTKPDSPVPLSMWIFCS 354
Db 78 npntcnennnggcdadakcteedsngskktctcctkpdspvplfdgifs 127

Search completed: August 8, 2001, 12:31:56
Job time: 164 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 8, 2001, 12:29:12 ; Search time 70.07 Seconds
(without alignments)
743.944 Million cell updates/sec

Title: US-09-500-376-2

Perfect score: 2062

Sequence: 1 AISVTMDNILSGFENEYDVI.....SNFLGISFLILMLILYSFI 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 segs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_16: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_unclassified: *
13: sp_vertebrate: *
14: sp_virus: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2043.5	99.1	651	5 Q25924	Q25924 plasmodium
2	2036.5	98.8	569	5 Q25983	Q25983 plasmodium
3	2030.5	98.5	569	5 Q25969	Q25969 plasmodium
4	2030.5	98.5	569	5 Q25974	Q25974 plasmodium
5	2030.5	98.5	569	5 Q25975	Q25975 plasmodium
6	2030.5	98.5	569	5 Q25977	Q25977 plasmodium
7	2030.5	98.5	569	5 Q25979	Q25979 plasmodium
8	2030.5	98.5	1694	5 Q9T2T5	Q9T2T5 plasmodium
9	2030.5	98.5	1694	5 Q9NHX1	Q9NHX1 plasmodium
10	2030.5	98.5	1704	5 Q9T2T4	Q9T2T4 plasmodium
11	2027	98.3	570	5 Q25968	Q25968 plasmodium
12	2027	98.3	570	5 Q9TYG2	Q9TYG2 plasmodium
13	2025	98.2	1720	5 Q25922	Q25922 plasmodium
14	2021	98.0	652	5 Q25923	Q25923 plasmodium
15	2017.5	97.8	569	5 Q25978	Q25978 plasmodium
16	2011.5	97.6	569	5 Q25967	Q25967 plasmodium
17	2011.5	97.6	569	5 Q25970	Q25970 plasmodium
18	2011.5	97.6	569	5 Q25980	Q25980 plasmodium
19	2011.5	97.6	569	5 Q25982	Q25982 plasmodium

20	1961	95.1	373	5	Q25724	plasmodium
21	1956	94.9	373	5	Q25721	plasmodium
22	1948	94.5	373	5	Q43995	plasmodium
23	1943.5	94.3	372	5	Q25725	plasmodium
24	1943	94.2	373	5	Q43996	plasmodium
25	1943	94.2	373	5	Q25723	plasmodium
26	1938.5	94.0	372	5	Q25726	plasmodium
27	1937.5	94.0	372	5	Q43997	plasmodium
28	1936	93.9	373	5	Q25722	plasmodium
29	1924	93.3	373	5	Q25727	plasmodium
30	1919	93.1	373	5	Q25728	plasmodium
31	1913.5	92.8	372	5	Q25717	plasmodium
32	1908.5	92.6	372	5	Q25718	plasmodium
33	1908.5	92.6	372	5	Q25719	plasmodium
34	1908.5	92.6	372	5	Q25720	plasmodium
35	1922.5	62.7	539	5	Q25972	plasmodium
36	1922.5	62.7	539	5	Q25981	plasmodium
37	1922.5	62.7	539	5	Q03999	plasmodium
38	1922.5	62.7	539	5	Q25966	plasmodium
39	1922.5	62.7	539	5	Q25976	plasmodium
40	1922.5	62.7	539	5	Q25984	plasmodium
41	1922.5	62.7	539	5	Q9TYG1	plasmodium
42	1922.5	62.7	539	5	Q25971	plasmodium
43	1922.5	62.7	539	5	Q25973	plasmodium
44	1922.5	62.7	539	5	Q9TVG8	plasmodium
45	845	41.0	1726	5	Q02569	plasmodium

ALIGNMENTS

RESULT	1
Q25924	
ID	Q25924
AC	Q25924; PRELIMINARY; PRT; 651 AA.
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE	MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=5833;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=RO-71;
RX	MEDLINE=92275047; PubMed=1592091;
RA	Olafsson P., Matile H., Certa U.;
RT	"Plasmodium falciparum: the repetitive MSA-1 surface protein of the RO-71 isolate is recognized by mouse antibody against the nonrepetitive repeat block of RO-33.";
RL	Exp. Parasitol. 74:381-389(1992).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=RO-71;
RX	MEDLINE=95354793; PubMed=7628566;
RA	Tolle R., Bujard H., Cooper J.A.;
RT	"Plasmodium falciparum: variations within the C-terminal region of merozoite surface antigen-1.";
RL	Exp. Parasitol. 81:47-54(1995).
DR	EMBL; Z35329; CAA84558.1;
DR	InterPro; IPR000561;
DR	Pfam; PF000008; EGF; 1.
KW	Merozoite.
FT	NON_TER
SQ	SEQUENCE 651 AA; 74134 MW; AA2137E699255150 CRC64;

Query Match 99.1%; Score 2043.5; DB 5; Length 651;
Best Local Similarity 99.5%; Pred. No. 2.2e-102;
Matches 392; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 AISVTMDNILSGFENEYDVIYKPLAGVYKSLKQIEKNIFTFNLDNLNLSRLKRRY 60
|||||

```
Db 259 AISVTMDNILSGFENEYDVYILKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKRRKY 318
Qy 61 FLDVLESDLMQFKHISNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 319 FLDVLESDLMQFKHISNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 378
Qy 121 YIEKVLAKYKDDLESIKKVIKEEKEFPSSPTTTPPSPAKTDEQKESKFLPFTLTNIETL 180
Db 379 YIEKVLAKYKDDLESIKKVIKEEKE-FPSSPTTTPPSPAKTDEQKESKFLPFTLTNIETL 437
Qy 181 YNNLVNKIDDYILNKLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
Db 438 YNNLVNKIDDYILNKLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIKK 497
Qy 241 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISOHCVKKQCPENSCGFR 300
Db 498 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISOHCVKKQCPENSCGFR 557
Qy 301 HLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKTCTCCTK 360
Db 558 HLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKTCTCCTK 617
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 618 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 651

RESULT 2
ID Q25983 PRELIMINARY; PRT; 569 AA.
AC Q25983;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13355; BAA02616.1; -
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64491 MW; 57A6B62FF72CE885 CRC64;
```

```
Query Match 98.8%; Score 2036.5; DB 5; Length 569;
Best Local Similarity 99.5%; Pred. No. 4.6e-102;
Matches 392; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKRRKY 60
Db 177 AISVTMDNILSGFENEYDVYILKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKRRKY 236
Qy 61 FLDVLESDLMQFKHISNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 296
Qy 121 YIEKVLAKYKDDLESIKKVIKEEKEFPSSPTTTPPSPAKTDEQKESKFLPFTLTNIETL 180
Db 297 YIEKVLAKYKDDLESIKKVIKEEKE-FPSSPTTTPPSPAKTDEQKESKFLPFTLTNIETL 355
Qy 181 YNNLVNKIDDYILNKLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
```

```
Db 356 YNNLVNKIDDYILNKLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIKK 415
Qy 241 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISOHCVKKQCPENSCGFR 300
Db 416 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISOHCVKKQCPENSCGFR 475
Qy 301 HLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKTCTCCTK 360
Db 476 HLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKTCTCCTK 535
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 569

RESULT 3
ID Q25969 PRELIMINARY; PRT; 569 AA.
AC Q25969;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13346; BAA02607.1; -
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match 98.5%; Score 2030.5; DB 5; Length 569;
Best Local Similarity 99.2%; Pred. No. 9.7e-102;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 AISVTMDNILSGFENEYDVYILKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKRRKY 60
Db 177 AISVTMDNILSGFENEYDVYILKPLAGVYRSLKKQIEKNFTFNLNLDILNSRLKRRKY 236
Qy 61 FLDVLESDLMQFKHISNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 296
Qy 121 YIEKVLAKYKDDLESIKKVIKEEKEFPSSPTTTPPSPAKTDEQKESKFLPFTLTNIETL 180
Db 297 YIEKVLAKYKDDLESIKKVIKEEKE-FPSSPTTTPPSPAKTDEQKESKFLPFTLTNIETL 355
Qy 181 YNNLVNKIDDYILNKLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIKK 240
Db 356 YNNLVNKIDDYILNKLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIKK 415
Qy 241 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISOHCVKKQCPENSCGFR 300
Db 416 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISOHCVKKQCPENSCGFR 475
Qy 301 HLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKTCTCCTK 360
Db 476 HLDERECKCLLNYKQBGDKCVENPNTCNENNGCGDADAKCTEEDSGSNGKKTCTCCTK 535
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
```

Db 536 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 569

RESULT 4

Q25974 ID Q25974 PRELIMINARY; PRT; 569 AA.
AC Q25974
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93295445; PubMed-8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13348; BAA02609.1; -;
DR InterPro; IPR000561; -;
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match 98.5%; Score 2030.5; DB 5; Length 569;
Best Local Similarity 99.2%; Pred. No. 9.7e-102;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AISVTMDNLGSGFENEYDVYLYKPLAGYRSLKQIEKNITFTNLDLNSRLKRRKY 60
DB 177 AISVTMDNLGSGFENEYDVYLYKPLAGYRSLKQIEKNITFTNLDLNSRLKRRKY 236
QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
DB 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 296
QY 121 YEKVLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFTNIETL 180
DB 297 YEKVLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFTNIETL 355
QY 181 YNNLVNKIDYLYNLKAKINDCNVKEDEAHVITPKLSDLKAIDDKIDLFKNHNDFEAIKK 240
DB 356 YNNLVNKIDYLYNLKAKINDCNVKEDEAHVITPKLSDLKAIDDKIDLFKNHNDFEAIKK 415
QY 241 LINDTCKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKQCPENSGCFR 300
DB 416 LINDTCKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKQCPENSGCFR 475
QY 301 HLDRECKCLLYNQKQGDCKVNPNTCNENGGCDAKCTEEDSGSGNGKKTTCCTCK 360
DB 476 HLDRECKCLLYNQKQGDCKVNPNTCNENGGCDAKCTEEDSGSGNGKKTTCCTCK 535
QY 361 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 394
DB 536 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 569

RESULT 5

Q25975 ID Q25975 PRELIMINARY; PRT; 569 AA.
AC Q25975
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.

OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93295445; PubMed-8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13349; BAA02610.1; -;
DR InterPro; IPR000561; -;
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64459 MW; ED4284B2867C9703 CRC64;

Query Match 98.5%; Score 2030.5; DB 5; Length 569;
Best Local Similarity 99.2%; Pred. No. 9.7e-102;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AISVTMDNLGSGFENEYDVYLYKPLAGYRSLKQIEKNITFTNLDLNSRLKRRKY 60
DB 177 AISVTMDNLGSGFENEYDVYLYKPLAGYRSLKQIEKNITFTNLDLNSRLKRRKY 236
QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
DB 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 296
QY 121 YEKVLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFTNIETL 180
DB 297 YEKVLAKYKDDLESIKKVIKEEKEFPSPPTPPSPAKTDEQKESKFLPFTNIETL 355
QY 181 YNNLVNKIDYLYNLKAKINDCNVKEDEAHVITPKLSDLKAIDDKIDLFKNHNDFEAIKK 240
DB 356 YNNLVNKIDYLYNLKAKINDCNVKEDEAHVITPKLSDLKAIDDKIDLFKNHNDFEAIKK 415
QY 241 LINDTCKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKQCPENSGCFR 300
DB 416 LINDTCKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKQCPENSGCFR 475
QY 301 HLDRECKCLLYNQKQGDCKVNPNTCNENGGCDAKCTEEDSGSGNGKKTTCCTCK 360
DB 476 HLDRECKCLLYNQKQGDCKVNPNTCNENGGCDAKCTEEDSGSGNGKKTTCCTCK 535
QY 361 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 394
DB 536 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 569

RESULT 6
Q25977

ID Q25977 PRELIMINARY; PRT; 569 AA.
AC Q25977
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93295445; PubMed-8515786;
RA Jongwutiwes S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13350; BAA02611.1; -;

```
DR InterPro: IPR000561; -.
DR Pfam: PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64485 MW; A26C45914AF01396 CRC64;

Query Match      98.5%; Score 2030.5; DB 5; Length 569;
Best Local Similarity 99.2%; Pred. No. 9.7e-102;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 60
Db 177 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 236

QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 296

QY 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 355

QY 181 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 240
Db 356 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 415

QY 241 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGGKQFQDMLNISQHCQVKKQCPENSGCFR 300
Db 416 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGGKQFQDMLNISQHCQVKKQCPENSGCFR 475

QY 301 HLDERECKCLLNYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSGNGKITCECTK 360
Db 476 HLDERECKCLLNYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSGNGKITCECTK 535

QY 361 PDSYPLFDGIFCSCSNFLGISFLLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSCSNFLGISFLLILMLILYSFI 569

RESULT 7
Q25979 ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HNI;
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062348; AAC72884.1;
DR InterPro: IPR000561; -.
KW Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match      98.5%; Score 2030.5; DB 5; Length 569;
Best Local Similarity 99.2%; Pred. No. 9.7e-102;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 60
Db 177 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 236

QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 296

QY 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 355

QY 181 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 240
Db 356 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 415

QY 241 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGGKQFQDMLNISQHCQVKKQCPENSGCFR 300
Db 416 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGGKQFQDMLNISQHCQVKKQCPENSGCFR 475

QY 301 HLDERECKCLLNYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSGNGKITCECTK 360
Db 476 HLDERECKCLLNYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSGNGKITCECTK 535

QY 361 PDSYPLFDGIFCSCSNFLGISFLLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSCSNFLGISFLLILMLILYSFI 569

RESULT 7
Q25979 ID Q25979 PRELIMINARY; PRT; 569 AA.
AC Q25979
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=93295445; PubMed=8515786;
RA Jungwutives S., Tanabe K., Kanbara H.;
RL "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates.";
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL: DJ3352; BAA02613.1; -.
DR InterPro: IPR000561; -.
DR Pfam: PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64457 MW; A2691FE569E80396 CRC64;

Query Match      98.5%; Score 2030.5; DB 5; Length 569;
Best Local Similarity 99.2%; Pred. No. 9.7e-102;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 60
Db 177 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 236

QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 237 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 296

QY 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 297 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 355

QY 181 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 240
Db 356 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 415

QY 241 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGGKQFQDMLNISQHCQVKKQCPENSGCFR 300
Db 416 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGGKQFQDMLNISQHCQVKKQCPENSGCFR 475

QY 301 HLDERECKCLLNYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSGNGKITCECTK 360
Db 476 HLDERECKCLLNYKQEGDKCVENPNTCNENGGCDADAKCTEEDSGSGNGKITCECTK 535

QY 361 PDSYPLFDGIFCSCSNFLGISFLLILMLILYSFI 394
Db 536 PDSYPLFDGIFCSCSNFLGISFLLILMLILYSFI 569

RESULT 8
Q9TZT5 ID Q9TZT5 PRELIMINARY; PRT; 1694 AA.
AC Q9TZT5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HNI;
RA Jiang G., Liu R.Z., Daubenberger C.A., Pluschke G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062348; AAC72884.1;
DR InterPro: IPR000561; -.
KW Pfam; PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192794 MW; 84CFC0E709F5673B CRC64;

Query Match      98.5%; Score 2030.5; DB 5; Length 1694;
Best Local Similarity 99.2%; Pred. No. 3e-101;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 60
Db 1302 AISVTMDNILSGFENEYDVIIYKPLAGVYSLKKQIEKNITFNLDILNSRLKRRKY 1361

QY 61 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 1362 FLDVLESDLMQFKHISSEYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 1421

QY 121 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 1422 YEKVLAKYKDDLESIKKVIKEEKFPSPPTTPPSPAKTDEQKESKFLPFLTNIETL 1480

QY 181 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 240
Db 1481 YNLVNVKIDYLLINLAKINDCNVEKDEAHVITKLSDLKAIDDKIDLPKNHNDFAIAKK 1540

QY 241 LINDTTKMDLGLKLLSTGLVQNFNTIISKLEGGKQFQDMLNISQHCQVKKQCPENSGCFR 300
```

```
|||||
Db 1541 LINDTKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 1600
Qy 301 HLDRECKCLLNTKQEGDKCVENPPTCNENNGGCCADAKCTEEDSGSNGKKITCECTK 360
Db 1601 HLDRECKCLLNTKQEGDKCVENPPTCNENNGGCCADAKCTEEDSGSNGKKITCECTK 1660
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 1661 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 1694

RESULT 9
Q9NHX1 PRELIMINARY; PRT; 1694 AA.
AC Q9NHX1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE MAJOR MEROZOITE SURFACE ANTIGEN.
GN GPI95.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Shan Z. X., Yu X. B., Li X. R., Ma C. L., Fang J. M.;
RT "Molecular cloning and sequence analysis of major merozoite surface
RT antigen(gpi95)gene of Plasmodium falciparum isolate FCCL/HN. ";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218248; AAF27526.1; -
DR InterPro; IPR000561; -
DR InterPro; IPR003247; -
DR Pfam; PF00008; EGF; 1.
DR ProDom; PD001527; -; 1.
KW Merozoite.
SQ SEQUENCE 1694 AA; 192766 MW; B51634A49E0F6728 CRC64;

Query Match 98.5%; Score 2030.5; DB 5; Length 1694;
Best Local Similarity 99.2%; Pred. No. 3e-101;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 AISVTMDNILSGFENEVDVYLLKPLAGYRSLLKQIEKNITFNLNLDILNSRLKKRY 60
Db 1302 AISVTMDNILSGFENEVDVYLLKPLAGYRSLLKQIEKNITFNLNLDILNSRLKKRY 1361
Qy 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 1362 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 1421
Qy 121 YEKVLAKYKDDLESIKKVIKEEKEFPSSPTTTPSPAKTDEQKESKFLPFLTNIETL 180
Db 1422 YEKVLAKYKDDLESIKKVIKEEKE -FPSSPTTTPSPAKTDEQKESKFLPFLTNIETL 1480
Qy 181 YNNLVNKIDYLLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 240
Db 1481 YNNLVNKIDYLLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 1540
Qy 241 LINDTKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 300
Db 1541 LINDTKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 1600
Qy 301 HLDRECKCLLNTKQEGDKCVENPPTCNENNGGCCADAKCTEEDSGSNGKKITCECTK 360
Db 1601 HLDRECKCLLNTKQEGDKCVENPPTCNENNGGCCADAKCTEEDSGSNGKKITCECTK 1660
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 1661 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 1694
```

```
RESULT 10
Q9TZT4 PRELIMINARY; PRT; 1704 AA.
AC Q9TZT4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR.
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HN2;
RA Jiang G., Liu R. Z., Daubenberger C. A., Pluschke G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF062349; AAC72885.1; -
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 1704 AA; 193762 MW; 385526D0DA56FD1D CRC64;

Query Match 98.5%; Score 2030.5; DB 5; Length 1704;
Best Local Similarity 99.2%; Pred. No. 3.1e-101;
Matches 391; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 AISVTMDNILSGFENEVDVYLLKPLAGYRSLLKQIEKNITFNLNLDILNSRLKKRY 60
Db 1312 AISVTMDNILSGFENEVDVYLLKPLAGYRSLLKQIEKNITFNLNLDILNSRLKKRY 1371
Qy 61 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 120
Db 1372 FLDVLESDLMQFKHISSEYIIEDSFLLNSEQNTLLKSYKIKESVENDIKFAQEGIS 1431
Qy 121 YEKVLAKYKDDLESIKKVIKEEKEFPSSPTTTPSPAKTDEQKESKFLPFLTNIETL 180
Db 1432 YEKVLAKYKDDLESIKKVIKEEKE -FPSSPTTTPSPAKTDEQKESKFLPFLTNIETL 1490
Qy 181 YNNLVNKIDYLLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 240
Db 1491 YNNLVNKIDYLLINLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK 1550
Qy 241 LINDTKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 300
Db 1551 LINDTKDMLGKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPNSGCFR 1610
Qy 301 HLDRECKCLLNTKQEGDKCVENPPTCNENNGGCCADAKCTEEDSGSNGKKITCECTK 360
Db 1611 HLDRECKCLLNTKQEGDKCVENPPTCNENNGGCCADAKCTEEDSGSNGKKITCECTK 1670
Qy 361 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 394
Db 1671 PDSYPLFDGIFCSSSNFLGISFLILMLILYSFI 1704

RESULT 11
Q25968 PRELIMINARY; PRT; 570 AA.
AC Q25968;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
```

Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates;
 Mol. Biochem. Parasitol. 59:95-100(1993).
 EMBL: D13343; BAA02606.1; -;
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 DR Merozoite; EGF-like domain.
 KW NON_TER 1 1
 FT SEQUENCE 570 AA; 64632 MW; 424BF553CCC2F2BE CRC64;

Query Match 98.3%; Score 2027; DB 5; Length 570;
 Best Local Similarity 98.7%; Pred. No. 1.5e-101;
 Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AISVTMDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLINSLRKKRY 60
 Db 177 AISVTMDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLINSLRKKRY 236
 QY 61 FLDVLESDLMQFHHISSNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
 Db 237 FLDVLESDLMQFHHISSNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 296
 QY 121 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 180
 Db 297 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 356
 QY 181 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIK 240
 Db 357 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIK 416
 QY 241 LINDTCKMGLKLLSTGLVQNFPTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 300
 Db 417 LINDTCKMGLKLLSTGLVQNFPTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 476
 QY 301 HLDRECKCLLNYKQEGDKCVENPNTCNENNGCGDADAKCTEEDSGSKKITCCTK 360
 Db 477 HLDRECKCLLNYKQEGDKCVENPNTCNENNGCGDADAKCTEEDSGSKKITCCTK 536
 QY 361 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 394
 Db 537 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 570

RESULT 12
 Q9TYG2 ID Q9TYG2 PRELIMINARY; PRT; 570 AA.
 AC Q9TYG2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88011243; PubMed=3079521;
 RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
 RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum".
 RL J. Mol. Biol. 195:273-287(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93295445; PubMed=8515786;
 RA Jongwattanas S., Tanabe K., Kanbara H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of Plasmodium falciparum from field isolates".
 RL Mol. Biochem. Parasitol. 59:95-100(1993).
 DR EMBL; D13343; BAA02604.1; -;

DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 DR Merozoite; EGF-like domain.
 FT NON_TER 1 1
 SQ SEQUENCE 570 AA; 64630 MW; 8674DEC89B2D662A CRC64;

Query Match 98.3%; Score 2027; DB 5; Length 570;
 Best Local Similarity 98.7%; Pred. No. 1.5e-101;
 Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AISVTMDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLINSLRKKRY 60
 Db 177 AISVTMDNLGSGFENEYDVYILKPLAGVYRSLLKQIEKNITFTNLDNLINSLRKKRY 236
 QY 61 FLDVLESDLMQFHHISSNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
 Db 237 FLDVLESDLMQFHHISSNEYIIEDSFLLNSEQNTLLSKYIKESVENDIKFAQEGIS 296
 QY 121 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 180
 Db 297 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFTNIETL 356
 QY 181 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIK 240
 Db 357 YNNLVNKIDDDYILNLKAKINDCNVEKDEAHVITKLSDLKAIDDKIDLFKNHNDFEAIK 416
 QY 241 LINDTCKMGLKLLSTGLVQNFPTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 300
 Db 417 LINDTCKMGLKLLSTGLVQNFPTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 476
 QY 301 HLDRECKCLLNYKQEGDKCVENPNTCNENNGCGDADAKCTEEDSGSKKITCCTK 360
 Db 477 HLDRECKCLLNYKQEGDKCVENPNTCNENNGCGDADAKCTEEDSGSKKITCCTK 536
 QY 361 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 394
 Db 537 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 570

RESULT 13
 Q25922 ID Q25922 PRELIMINARY; PRT; 1720 AA.
 AC Q25922;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
 DE PRECURSOR OF THE MAJOR MEROZOITE SURFACE ANTIGENS.
 OS Plasmodium falciparum (isolate NF54).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5843;
 RN [1]
 RP SEQUENCE OF 1069-1720 FROM N.A.
 RC STRAIN=NF54;
 RA Tolle R., Bujard H., Cooper J.A.;
 RL Exp. Parasitol. 0:0-0(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NF54;
 RA Tolle R.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NF54;
 RX MEDLINE=96123395; PubMed=8577332;
 RA Pan W., Tolle R., Bujard H.;
 RT "A direct and rapid sequencing strategy for the Plasmodium falciparum antigen gene gp190/MSA1".
 RL Mol. Biochem. Parasitol. 73:241-244(1995).
 DR EMBL; Z35327; CAA84556.1; -;
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 KW Signal; Merozoite.

```

FT SIGNAL 1 19 POTENTIAL.
SQ SEQUENCE 1720 AA; 195725 MW; 717B2FD1E637A8A3 CRC64;

Query Match 98.28; Score 2025; DB 5; Length 1720;
Best Local Similarity 98.7%; Pred. No. 6.1e-101;
Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGPFENEYDVYILKPLAGYVRSLLKQIEKNITFTNLDILNSRLKRRKY 60
Db 1327 AISTVMDNLSGPFENEYDVYILKPLAGYVRSLLKQIEKNITFTNLDILNSRLKRRKY 1386
Qy 61 FLDVLESQDMQFKHSSNEXYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 1387 FLDVLESQDMQFKHSSNEXYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 1446
Qy 121 YEKVLAKYKDDLESIRKVIKEEKFPSSPTTTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 1447 YEKVLAKYKDDLESIRKVIKEEKFPSSPTTTPPSPAKTDEQKESKFLPFLTNIETL 1506
Qy 181 YNNLVNKIDDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDPEAIKK 240
Db 1507 YNNLVNKIDDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDPEAIKK 1566
Qy 241 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 300
Db 1567 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 1626
Qy 301 HLDRECKCLLYNQEGDKCVENPNPTCNENNGGCDADAKTEEDSGSNGKKITCECTK 360
Db 1627 HLDRECKCLLYNQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSNGKKITCECTK 1686
Qy 361 PDSYPLFDGIFGICSSSNFLGISFLILMLILYSFI 394
Db 1687 PDSYPLFDGIFGICSSSNFLGISFLILMLILYSFI 1720

RESULT 14
Q25923 PRELIMINARY; PRT; 652 AA.
ID Q25923 AC Q25923
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE MEROZOITE SURFACE ANTIGEN 1 (FRAGMENT).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 1-298 FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=88142993; PubMed=2449612;
RA Peterson G.M., Coppel R.L., McIntyre P., Langford C.J., Woodrow G.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "Variation in the precursor to the major merozoite surface antigens of
RT Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 27:291-302(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RA Tolle R.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FCH5/NF7;
RX MEDLINE=9334793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: Variations within the C-terminal region of
RT merozoite surface antigen-1."
RL Exp. Parasitol. 81:47-54(1995).
DR EMBL; 233328; CAA84557.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.

```

```

KW Merozoite.
FT NON_TER 1
SQ SEQUENCE 652 AA; 74292 MW; 2B6A87737B490A62 CRC64;

Query Match 98.0%; Score 2021; DB 5; Length 652;
Best Local Similarity 98.5%; Pred. No. 3.6e-101;
Matches 388; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGPFENEYDVYILKPLAGYVRSLLKQIEKNITFTNLDILNSRLKRRKY 60
Db 259 AISTVMDNLSGPFENEYDVYILKPLAGYVRSLLKQIEKNITFTNLDILNSRLKRRKY 318
Qy 61 FLDVLESQDMQFKHSSNEXYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 120
Db 319 FLDVLESQDMQFKHSSNEXYIIEDSFKLLNSEQNTLLSKYIKESVENDIKFAQEGIS 378
Qy 121 YEKVLAKYKDDLESIRKVIKEEKFPSSPTTTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 379 YEKVLAKYKDDLESIRKVIKEEKFPSSPTTTPPSPAKTDEQKESKFLPFLTNIETL 438
Qy 181 YNNLVNKIDDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDPEAIKK 240
Db 439 YNNLVNKIDDDYLLNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDPEAIKK 498
Qy 241 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 300
Db 499 LINDTTRKMDLGLKLLSTGLVQNFNTIISKLEGFQDMLNISQHCVKKQCPENSGCFR 558
Qy 301 HLDRECKCLLYNQEGDKCVENPNPTCNENNGGCDADAKTEEDSGSNGKKITCECTK 360
Db 559 HLDRECKCLLYNQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSNGKKITCECTK 618
Qy 361 PDSYPLFDGIFGICSSSNFLGISFLILMLILYSFI 394
Db 619 PDSYPLFDGIFGICSSSNFLGISFLILMLILYSFI 652

RESULT 15
Q25978 PRELIMINARY; PRT; 569 AA.
ID Q25978 AC Q25978
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93295445; PubMed=8515786;
RA Jongwutives S., Tanabe K., Kanbara H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSP1) of Plasmodium falciparum from
RT field isolates."
RL Mol. Biochem. Parasitol. 59:95-100(1993).
DR EMBL; D13351; BAA02612.1; -.
DR InterPro; IPR000561; -.
DR Pfam; PF00008; EGF; 1.
KW Merozoite; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 569 AA; 64536 MW; B8B9B63EECB1DA51 CRC64;

Query Match 97.8%; Score 2017.5; DB 5; Length 569;
Best Local Similarity 98.7%; Pred. No. 4.8e-101;
Matches 389; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 AISTVMDNLSGPFENEYDVYILKPLAGYVRSLLKQIEKNITFTNLDILNSRLKRRKY 60
Db 177 AISTVMDNLSGPFENEYDVYILKPLAGYVRSLLKQIEKNITFTNLDILNSRLKRRKY 236

```

Qy	61	FLDVLESDLMQFKHISSNEVYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS	120
Db	237	FLDVLESDLMQFKHISSNEVYIIEDSFKLLNSEQNTLLKSYKIKESVENDIKFAQEGIS	296
Qy	121	YIEKVLAKYKDDLESIKKVIKEEKEPPSPPTTPPSPAKTDEQKESKFLPFLTNIETL	180
Db	297	YIEKVLAKYKDDLESIKKVIKEEKE-PPSSPPTTPPSPAKTDEQKESKFLPFLTNIETL	355
Qy	181	YNNLVNKIDDOYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK	240
Db	356	YNNLVNKIDDOYLINLKAINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNHNDFEAIKK	415
Qy	241	LINDDTKMDLGLSTGLVONPNTIISKLIIEGKFODMLNISOHQCVKKOCPENSGCFR	300
Db	416	LINDDTKMDLGLSTGLVONPNTIISKLIIEGKFODMLNISOHQCVKKOCPENSGCFR	475
Qy	301	HLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADAKTEEDSGSNGKKITCECTK	360
Db	476	HLDERECKCLLNYKQEGDKCVENPNPTCNENNGGCDADATCTEEDSGSSRKKITCECTK	535
Qy	361	PDYPLFDGIFCSSNFGISFLLILMLILYSFI	394
Db	536	PDYPLFDGIFCSSNFGISFLLILMLILYSFI	569

Search completed: August 8, 2001, 12:33:42
Job time: 270 sec


```
SQ SEQUENCE 1701 AA; 193719 MW; 3920B75E73D38552 CRC64;

Query Match 99.1%; Score 2033; DB 1; Length 1701;
Best Local Similarity 99.2%; Pred. No. 2.7e-97;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGFENEYDVYILKPLAGYRSLKQIEKNITFNLDNLINSLRLKRRKY 60
Dd 1308 AISTVMDNLSGFENEYDVYILKPLAGYRSLKQIEKNITFNLDNLINSLRLKRRKY 1367

Qy 61 FLDVLESLDMQFKHISSEYIIEFSKLLNSEQNKILLKSYKIKESVENDIKFAQEGIS 120
Dd 1368 FLDVLESLDMQFKHISSEYIIEFSKLLNSEQNKILLKSYKIKESVENDIKFAQEGIS 1427

Qy 121 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITEL 180
Dd 1428 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITEL 1487

Qy 181 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Dd 1488 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy 241 LINDDTKMDLGLKLLSTGLVQNPNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 300
Dd 1548 LINDDTKMDLGLKLLSTGLVQNPNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 1607

Qy 301 HLDRECKCLLNKQEGDKCEENPNPTCNENGGCDADATCTEEDSGSRKKITCECTK 360
Dd 1608 HLDRECKCLLNKQEGDKCEENPNPTCNENGGCDADATCTEEDSGSRKKITCECTK 1667

Qy 361 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 394
Dd 1668 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 1701

RESULT 2
MSPL PLAFM STANDARD; PRT; 1701 AA.
AC P08569;
DT 01-AUG-1988 (Rel. 08, Created)
DT 30-AUG-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70153;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88011243; PubMed=3079521;
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum.";
RL J. Mol. Biol. 195:273-287(1987).
RN [2]
RP REVISIONS TO 1403; 1569 AND 1629.
RA Tanabe K.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-115 FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; X05624; CAA29112.1; -
PRL; A28668; A28668.
PIR; B25120; B25120.
InterPro; IPR000561; -
Pfam; PF00008; EGF; 1.
Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein; Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 110 110 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 470 470 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 607 607 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 899 899 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 991 991 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1701 AA; 193768 MW; 3FC2EC59AF96EA98 CRC64;

Query Match 99.1%; Score 2033; DB 1; Length 1701;
Best Local Similarity 99.2%; Pred. No. 2.7e-97;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGFENEYDVYILKPLAGYRSLKQIEKNITFNLDNLINSLRLKRRKY 60
Dd 1308 AISTVMDNLSGFENEYDVYILKPLAGYRSLKQIEKNITFNLDNLINSLRLKRRKY 1367

Qy 61 FLDVLESLDMQFKHISSEYIIEFSKLLNSEQNKILLKSYKIKESVENDIKFAQEGIS 120
Dd 1368 FLDVLESLDMQFKHISSEYIIEFSKLLNSEQNKILLKSYKIKESVENDIKFAQEGIS 1427

Qy 121 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITEL 180
Dd 1428 YEKVLAKYKDDLESIKKVIKEEKEKPPSPPTPPSPAKTDEQKESKFLPFLTNITEL 1487

Qy 181 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Dd 1488 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1547

Qy 241 LINDDTKMDLGLKLLSTGLVQNPNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 300
Dd 1548 LINDDTKMDLGLKLLSTGLVQNPNTIISKLEGFQDMLNISQHCYKQCPENSGCFR 1607

Qy 301 HLDRECKCLLNKQEGDKCEENPNPTCNENGGCDADATCTEEDSGSRKKITCECTK 360
Dd 1608 HLDRECKCLLNKQEGDKCEENPNPTCNENGGCDADATCTEEDSGSRKKITCECTK 1667

Qy 361 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 394
Dd 1668 PDSYPLFDGIFCSSNFGISFLILMLILYSFI 1701

RESULT 3
MSPL PLAFM STANDARD; PRT; 1726 AA.
AC P04934;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
```

01-OCT-1996 (Rel. 34, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
 DE (PMWSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Camp / Malaysia).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5835;
 RN [1]
 RP SEQUENCE OF 1-1103 FROM N.A.
 RX MEDLINE=86205236; PubMed=3517809;
 RA Weber J.L., Leininger W.M., Lyon J.A.;
 RT "Variation in the gene encoding a major merozoite surface antigen of
 the human malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 14:3311-3323(1986).
 RN [2]
 RP SEQUENCE OF 1104-1726 FROM N.A.
 RX MEDLINE=88143999; PubMed=3278296;
 RA Weber J.L., Sim B.K.L., Lyon J.A., Wolff R.;
 RT "Merozoite surface protein sequence from the Camp strain of the human
 malaria parasite Plasmodium falciparum.";
 RL Nucleic Acids Res. 16:1206-1206(1988).
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -!- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03831; CAA27446.1; -;
 DR PIR; A23386; SAZQGM.
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MERZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1726 AA; 196197 MW; D8AD45FA352BCF3 CRC64;

Query Match 97.88; Score 2006; DB 1; Length 1726;
 Best Local Similarity 98.0%; Pred. No. 6.6e-96;
 Matches 386; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 AISTVMDNLGSGFENEVDVYVLLKPLAGYVSLKQIEKNILFTNLNLDILNSLRKKRY 60
 |||||||
 Db 1333 AISTVMDNLGSGFENEVDVYVLLKPLAGYVSLKQIEKNILFTNLNLDILNSLRKKRY 1392
 |||||||
 QY 61 FLDVLESLDMQFKHISSEYIIEFSLNSEQNLKSKYIKESVENDIKFAQEGIS 120
 |||||||
 Db 1393 FLDVLESLDMQFKHISSEYIIEFSLNSEQNLKSKYIKESVENDIKFAQEGIS 1452
 |||||||
 QY 121 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTTTPSPAKTDEQKESKFLPFTNIETL 180
 |||||||

Db 1453 YEKVLAKYKDDLESIRKVIKEEKEKFPSPPTTTPSPAKTDEQKESKFLPFTNIETL 1512
 QY 181 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKITKSLDKAIDDKIDLFKTNDFEAIKK 240
 |||||||
 Db 1513 YNNLVNKIDDYLLNLKAKINDCNVKEDEAHVKITKSLDKAIDDKIDLFKTNDFEAIKK 1572
 |||||||
 QY 241 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEIGKFDMLNISQHCVKKQCPENSGCFR 300
 |||||||
 Db 1573 LINDTTRKMDLGLKLLSTGLVQIFPNTIISKLEIGKFDMLNISQHCVKKQCPENSGCFR 1632
 |||||||
 QY 301 HLDRECKCLLNYKQEGDKCEENPNTCNENNGGCDADATCTEEDSGSRKKITCECTK 360
 |||||||
 Db 1633 HLDRECKCLLNYKQEGDKCEENPNTCNENNGGCDADARCTEEDSGSGNGKITCECTK 1692
 |||||||
 QY 361 PDSVPLFDGIFCSSLNFGISFLILMLILYSFI 394
 |||||||
 Db 1693 PDSVPLFDGIFCSSLNFGISFLILMLILYSFI 1726
 |||||||

RESULT 4
 MSPI_PLAAPP
 ID MSP1_PLAPP STANDARD; PRT; 1726 AA.
 AC P50435;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MERZOITE SURFACE PROTEIN 1 PRECURSOR (MERZOITE SURFACE ANTIGENS)
 DE (PMWSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Palo Alto / Uganda).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57270;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89005525; PubMed=3049134;
 RA Chang S.P., Kramer K.J., Yamaga K.M., Kato A., Case S.E.,
 RA Siddiqui W.A.;
 RT "Plasmodium falciparum: gene structure and hydrophathy profile of the
 major merozoite surface antigen (gpi95) of the Uganda-Palo Alto
 isolate";
 RL Exp. Parasitol. 67:1-11(1988).
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -!- PTM: MERZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MERZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M37213; AAA29611.1; -;
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1726 MERZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 567 567 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 638 638 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 827 827 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 924 924 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 944 944 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 990 990 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1016 1016 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1221 1221 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1613 1613 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1726 AA; 196174 MW; 5B59CEFA2F9A026 CRC64;

Query Match 97.6%; Score 2003; DB 1; Length 1726;
Best Local Similarity 97.7%; Pred. No. 9.4e-96;
Matches 385; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AISTVMDNLSGFENEVDVYILKPLAGVYSLKQIEKNIITFNLDILNSLRKRY 60
Db 1333 AISTVMDNLSGFENEVDVYILKPLAGVYSLKQIEKNIITFNLDILNSLRKRY 1392
Qy 61 FLDVLESLDMQFKHISSEYIIESDFKLLNSEQNTLLKSYIKESVENDIKFAQEGIS 120
Db 1393 FLDVLESLDMQFKHISSEYIIESDFKLLNSEQNTLLKSYIKESVENDIKFAQEGIS 1452
Qy 121 YEKVLAKYKDDLESIRKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 1453 YEKVLAKYKDDLESIRKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 1512
Qy 181 YNNLVNKIDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 1513 YNNLVNKIDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1572
Qy 241 LINDTKMDLGLKLLSTGLVQIFPNTIISKLEGFODMLNISQHCVKQCPENSCGFR 300
Db 1573 LINDTKMDLGLKLLSTGLVQIFPNTIISKLEGFODMLNISQHCVKQCPENSCGFR 1632
Qy 301 HLDRECKCLLYNQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCETCK 360
Db 1633 HLDRECKCLLYNQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCETCK 1692
Qy 361 PDSYPLPDGIFCSSLNPLGISFLLILMLILYSFI 394
Db 1693 PDSYPLPDGIFCSSLNPLGISFLLILMLILYSFI 1726

RESULT 5
MSPL_PLAF3 STANDARD; PRT; 1682 AA.
AC P19596; Q25921;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate ro-33 / Ghana).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5834;
RN [1]
RP SEQUENCE OF 1-1061 FROM N.A.
RX MEDLINE=8816657; PubMed=3327688;
RA Certa U., Rotmann D., Matile H., Reber-Liske R.;
RT "A naturally occurring gene encoding the major surface antigen
precursor p190 of plasmodium falciparum lacks tripeptide repeats.";
RL EMBO J. 6:4137-4142(1987).
RN [2]
RP SEQUENCE OF 1032-1682 FROM N.A.
RX MEDLINE=95354793; PubMed=7628566;
RA Tolle R., Bujard H., Cooper J.A.;
RT "Plasmodium falciparum: variations within the C-terminal region of
merozoite surface antigen-1.";
RL Exp. Parasitol. 81:47-54(1995).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M35727; AAA29715.1; -;
DR EMBL; Y00087; CAA68280.1; -;
DR EMBL; Z35326; CAA84555.1; -;
DR PIR; S06286; S06286;
DR InterPro; IPR000561; -;
DR Pfam; PF00008; EGF; 1;
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 13 POTENTIAL.
FT CHAIN 20 1682 MEROZOITE SURFACE PROTEIN 1.
FT TRANSMEM 1666 1682 MEMBRANE ANCHOR (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 785 785 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 901 901 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 947 947 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1178 1178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1569 1569 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1682 AA; 192462 MW; C82AJE159948CAD6 CRC64;

Query Match 96.5%; Score 1980.5; DB 1; Length 1682;
Best Local Similarity 97.2%; Pred. No. 1.3e-94;
Matches 383; Conservative 2; Mismatches 8; Indels 1; Gaps 1;

Qy 1 AISTVMDNLSGFENEVDVYILKPLAGVYSLKQIEKNIITFNLDILNSLRKRY 60
Db 1290 AISTVMDNLSGFENEVDVYILKPLAGVYSLKQIEKNIITFNLDILNSLRKRY 1349
Qy 61 FLDVLESLDMQFKHISSEYIIESDFKLLNSEQNTLLKSYIKESVENDIKFAQEGIS 120
Db 1350 FLDVLESLDMQFKHISSEYIIESDFKLLNSEQNTLLKSYIKESVENDIKFAQEGIS 1409
Qy 121 YEKVLAKYKDDLESIRKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 180
Db 1410 YEKVLAKYKDDLESIRKVIKEEKFPSPPTPPSPAKTDEQKESKFLPFLTNIETL 1468
Qy 181 YNNLVNKIDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 240
Db 1469 YNNLVNKIDDYILNLKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLFKNTNDFEAIKK 1528
Qy 241 LINDTKMDLGLKLLSTGLVQIFPNTIISKLEGFODMLNISQHCVKQCPENSCGFR 300
Db 1529 LINDTKMDLGLKLLSTGLVQIFPNTIISKLEGFODMLNISQHCVKQCPENSCGFR 1588
Qy 301 HLDRECKCLLYNQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCETCK 360
Db 1589 HLDRECKCLLYNQEGDKCEENPNTCNENNGCGDADATCTEEDSGSRKKITCETCK 1648
Qy 361 PDSYPLPDGIFCSSLNPLGISFLLILMLILYSFI 394
Db 1649 PDSYPLPDGIFCSSLNPLGISFLLILMLILYSFI 1682

RESULT 6
MSPL_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).

GN MSP-1.
 OS Plasmodium falciparum (isolate K1 / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stunnenberg H., Bujard H.;
 RA "Polymorphism of the precursor for the major surface antigens of
 RT Plasmodium falciparum merozoites: studies at the genetic level.";
 RL EMBO J. 4:3823-3829(1985).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA Pan W., Tolle R., Bujard H.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03371; CAA27070.1; -;
 DR PIR; A25120; SAZOK1.
 DR InterPro; IPR000561; -;
 DR Pfam; PF00008; EGF; 1.
 DR Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
 FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
 FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 955 955 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1630 AA; 187289 MW; ADHDEC3CE0A46322 CRC64;
 Query Match 61.1%; Score 1254.5; DB 1; Length 1630;
 Best Local Similarity 61.2%; Pred. No. 2.1e-57;
 Matches 240; Conservative 54; Mismatches 73; Indels 25; Gaps 3;
 QY 6 MDNLSGFENEYDVIYKPLAGVYRSLSKQIEKNIITFNLMNDILNSRLKRRKYFDLVL 65
 DB IDNLSKIENEYEVLYLPLAGVYRSLSKQLENNVFNVVKDILNSRFNKRNFKNVL 1320
 QY 66 ESDLMOQFKHSSNYIYEDSKFLNSQKNLLSKYIKESVENDIKFAQEGISYYEKV 125
 DB ESLLPYKDLTSSNYVVRDYPKFLNKEKROKFLSSYNIKIDSDTDNFANDVLGYKIL 1380
 QY 126 LAKYKDDLESIKYKIEKEKFPSSPTTPSPAKTDQEKESKFLPLNFIETLYNNLV 185
 DB SEKYSDLSDSIKKYI-----NDKQGENEKYLPFLNNFIETLYKTVN 1420
 QY 186 NKIDYDYLNLKAKINDCNVEKDEARVKITKLSDLKADDDKIDLPKNTNDFEAKKLNDD 245

RESULT 7

MSPI_PLAFW

ID MSPI_PLAFW

AC P04933;

DT 13-AUG-1987 (Rel. 05, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)

DE (PMMSA) (P195).

GN MSP-1.

OS Plasmodium falciparum (isolate Wellcome).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5848;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=86014355; PubMed=2995820;

RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,

RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,

RA Freeman R.R.;

RT "Primary structure of the precursor to the three major surface

RT antigens of Plasmodium falciparum merozoites.";

RL Nature 317:270-273(1985).

RN [2]

RP REVISIONS

RA Holder A.A.;

RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR

CC (POTENTIAL).

CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42

CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X02919; CAA26676.1; -;

CC PIR; A24594; A24594.

CC InterPro; IPR000561; -;

CC Pfam; PF00008; EGF; 1.

CC Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;

KW Transmembrane; GPI-anchor.

FT SIGNAL 1 19 POTENTIAL.

FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.

FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1639 AA; 187618 MW; 2C25586616C87F6E CRC64;

Query Match 61.1%; Score 1254.5; DB 1; Length 1639;
 Best Local Similarity 61.2%; Pred. No. 2.1e-57;
 Matches 240; Conservative 54; Mismatches 73; Indels 25; Gaps 3;

Qy 6 MNILSGFENEYDIYKPLAGVYRSKQIEKNIIITFNILNLDILNSRLKRRKYFLDVL 65
 Db 1270 IDNLSKIENEYEVLYKPLAGVYRSKQLENNVTFNVKDLNSRKNRKNFNVL 1329

Qy 66 ESDLMQFKHSSNYIYEDSKLNSQKILLKSYKIESVENDIKFAQEGISYYEKV 125
 Db 1330 ESDLPYKDLTSSNYVYKDYKFLNKKRDKFLSSYVYIKSDIDTFANDVLGYKIL 1389

Qy 126 LAKYKDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKESKFLPLNIETLYNNLV 185
 Db 1390 SEKYSKDSLSKKYI-----NDKQGENEKYLPFLNIETLYKTVN 1429

Qy 186 NKIDYDLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKFNTNDEFAIKKLIND 245
 Db 1430 DKIDFVHLEAKVLYTYEKSNNVEVKELNYLKTIDQKLPKNNFVGIADLSTDY 1489

Qy 246 TKDKMLKLLSTGLVQIFPN---TIISKLEGRFQDMLNISOHCYKQCPENSGCFRHL 302
 Db 1490 NHNNLTKFLSTGMV---FENLAKITVNLNLDGNIQGMNLNISOHCYKQCPENSGCFRHL 1547

Qy 303 DERECCLNLYKQEGDKCENPNPTCNENNGGCDADATCTEEDSGSRKKITCECKPD 362
 Db 1548 DERECCLNLYKQEGDKCENPNPTCNENNGGCDADAKTEEDSGSGNKKITCECKPD 1607

Qy 363 SYPLFDGIFGSSNFLGISFLILMLILYSFI 394
 Db 1608 SYPLFDGIFGSSNFLGISFLILMLILYSFI 1639

RESULT 8
 MSPI_PLAYO STANDARD; PRT; 1772 AA.
 ID MSPI_PLAYO
 AC P13828;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMMSA) (230 KDA).
 GN MSP-1.
 OS Plasmodium berghei yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID:5862;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90205979; PubMed=2320061;
 RA Lewis A.P.;
 RT "Sequence analysis upstream of the gene encoding the precursor to the
 RT major merozoite surface antigens of Plasmodium yoelii.";
 RL Mol. Biochem. Parasitol. 39:285-288(1990).
 RN [2]
 RP SEQUENCE OF 1093-1772 FROM N.A.
 RC STRAIN=17XL;
 RX MEDLINE=88124889; PubMed=2448778;
 RA Burns J.M. Jr., Daly T.M., Vaidya A.B., Long C.A.;
 RT "The 3' portion of the gene for a plasmodium yoelii merozoite surface
 RT antigen encodes the epitope recognized by a protective monoclonal
 RT antibody.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:602-606(1988).
 CC -1- SURCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF

CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J03612; AAA29762.1; -
 CC EMBL; J04668; AAA29702.1; -
 CC PIR; A28121; A28121.
 CC PIR; A45532; A45532.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 1772 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 406 406 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 829 829 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1018 1018 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1090 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1408 1408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1446 1446 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1541 1541 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1629 1629 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1680 1680 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1521 1521 L -> V (IN REF. 2).
 SQ SEQUENCE 1772 AA; 197230 MW; 9A6291658EB0F45D CRC64;

Query Match 28.9%; Score 593.5; DB 1; Length 1772;
 Best Local Similarity 32.5%; Pred. No. 1.8e-23;
 Matches 128; Conservative 96; Mismatches 137; Indels 33; Gaps 8;

Qy 8 NILSGFENEYDIYKPLAGVYRSKQIEKNIIITFNILNLDILNSRLKRRKYFLDVL 67
 Db 1401 DILSEFTNSLYVTYTKRGLSTYSLKHKMLREFSTIKEDMTNGLNKSQRNDFLEVL 1460

Qy 68 DLMOFKHISSNEYIIDSFKLLNSEQKNILLKSYKIESVENDIKFAQEGISYYEKVLA 127
 Db 1461 ELDFKDLSTNKVIRNPYQLDNDKDKQIVNLKATKGINEDIETTDGINKFNKWE 1520

Qy 128 KYDDLESIKKVIKEKEKFPSPPTTPSPAKTDEQKE--SKFLPFLNIETLYNNLV 185
 Db 1521 LYTQLAARVEQIATIE-----AETNDTNKEKKYIPILEDLKLGLYETVI 1566

Qy 186 NKIDYDLINKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLF---KNTNDFEATKKL 241
 Db 1567 GQAEVSEELQNRDLNLYKNEKAEPILTKNLEKYIQIDKLEDFEVEHAENKHIASL--A 1624

Qy 242 INDTRKMDLGLKLLSTGLVQIFPNTIISKLEGRFQDMLNIS-QHCYV-KKQCPENSGCF 299
 Db 1625 LNNLKSGLVGES-----KKILAKMLNMDGMDLLGVDPKHVCVDRDPKNAAGCF 1676

Qy 300 RHLDERECKLLNYKQ-EGDKCEENPNPTCNENNGGCDADATCTEEDSGSRKKITCEC 358
 Db 1677 RDDNGTEWRCLLYKKGEGTCVENNPTCDINNGGCDPTASCQNAESTENSKKIITC 1736

Qy 359 TKPDSYPLFDGIFGSSNFLGISFLILMLILYS 392
 Db 1737 KEPTPNAYEGVFCSSSFMGLSILLIITLIVFN 1770

RESULT 9
 YL17_CAEEL STANDARD; PRT; 1130 AA.
 ID YL17_CAEEL
 AC Q11102;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL 131.5 KDA PROTEIN C02F12.7 IN CHROMOSOME X.
GN C02F12.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Miller N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: WEAK, TO MYOSIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U41545; AAA83190.1; -;
DR Wormpep; C02F12.7; CE03901.
DR Hypothetical protein; Coiled coil.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 121 779 COILED COIL (POTENTIAL).
FT DOMAIN 805 1061 COILED COIL (POTENTIAL).
SQ SEQUENCE 1130 AA; 131485 MW; B0FD2EFE3D99FB09 CRC64;

Query Match 7.2%; Score 147; DB 1; Length 1130;
Best Local Similarity 23.6%; Pred. No. 0.9;
Matches 77; Conservative 64; Mismatches 129; Indels 56; Gaps 14;

Qy 31 SLKQIEKNIITFNLNDILNLRKK-----RYFLDYLESILMOPKHISN--EY 80
Db 596 NLRKOLEKE-ISHTEDNRLLHENTQKLEAHKETHETVTVLEAIDQFKSAFENEQY 654

Qy 81 IEDSFKLLNSEQN--ILKSYKYIKSVENDIKFAQBGISYKVLAKYKDDLESIKV 139
Db 655 GKESAKIRELEAQNKLTLSEMKVHVAEENLEAFTSDKNLLELESKNK-NIEHLKQE 713

Qy 140 IKEKEKFPSPPTPPSPAKTDSQKESKFLPFLNIETLYNNLVNKKIDDYLINLAKI 199
Db 714 IALNLEKI-----STKETEKDSELEKTIQALQIDNSKSDQIEKLHLRVNDML 761

Qy 200 NDCNVKDEAHVKITKLSDLKAIIDKIDLFKNTNDFEAIKKL--INDTKKMDLGLKLLST 257
Db 762 DQMTIKDEL---VKNKEEIKTISAKTAQLLESNTVSETKLASVTEEREKE----- 810

Qy 258 GLVQIFPNTIISKIEGKFQDMLNISQ-----HOCVKKQCPENSGCFRHLDE-----REE 307
Db 811 --IQSF-QTQISE---KDNVLTFAERINELETCLAREVELTGMRTKLDMDTQQLNEE 863

Qy 308 CKCLLYNKQEGDKCEPNPNTCENN 333
Db 864 TTVLVFONSIQEKIDEX-EATINEMN 888

RESULT 10
Y109_YEAST STANDARD; PRT; 1679 AA.
AC P40457;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHETICAL 195.1 KDA PROTEIN IN DNA43-UBI1 INTERGENIC REGION.
GN Y1L149C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z38059; CAA86129.1; -;
DR PIR; S48385; S48385.
DR SGD; S0001411; MLP2.
KW Hypothetical protein.
SQ SEQUENCE 1679 AA; 195141 MW; 298950CC52202D8F CRC64;

Query Match 7.2%; Score 147; DB 1; Length 1679;
Best Local Similarity 22.7%; Pred. No. 1.4;
Matches 105; Conservative 70; Mismatches 150; Indels 138; Gaps 22;

Qy 3 SVTWDNLGSGFENYDYIYKPLAGVYRSKKQ---TEKNIITFNLNLDILNLRKKR 59
Db 1079 SVLIEKVDDTAAANGDKDHLK-LVSLFSLNLRHNSLETKLTKCKREL-----AFVKOKN 1132

Qy 60 YFLDYLESILMOPKHISNEY-----IIEDSFKLLNSE--QKNILLKSYKYIKSVEN--- 110
Db 1133 DSLEKTIINDLQRTOTLTSEKYYQCSAVIIDEFKDTKEVTQVNLKNNAILQKSLKNVTE 1192

Qy 111 -----DIKPAQGISYKVLAKYKDDLE-SIKKVKKEEKFPSPPTTPSPAKTD 162
Db 1193 KNREIYKOLNDROQEISFLORDLIOTKEQVINSKILVYESEMEQCKQRYQDLSQOQKD 1252

Qy 163 BOKKESFLPFLNIETLYNNLVNKKIDDYLINLAKI-----NDCNVKDEAHV 211
Db 1253 AQKXD-----IEKLTNEISD-----LKGKLSAENANADLENKFNRLKQAHE 1295

Qy 212 KI-----TKLSDLKAIIDDKI--DL-FKNTNDFEAIKKL-----INDTKK 248
Db 1296 KLDASKQQAALTNELNELKAIKDKLEQDLHFENAKVIDLDTKLKAHELOQSDVSRDHEK 1355

Qy 249 DMLGKLLS-----TGLVQIF-----PNTII---SKLIEGKFQDML 280
Db 1356 DTYRTLMEETESLAKKELQIFKTANSSSDAFEKLVNMEKEKDRIDERTKEFEKQLQETL 1415

Qy 281 NISQH-----QCVKKQCPQ--NSGCFRHLDERP-----CKCLLYNK 315
Db 1416 NKSTSSAEAYSKDITLKLKEWLKEYEDELRLRIKAEENLKKIRLPSEERIQIKSRK 1475

Qy 316 QEGD-----KCEENPNTCENNNGCCDADATCTEEDSGSSRRK 353
Db 1476 EEELEEFKRLKENAGSLTFLDNKSGGDA---BEELWNPSK 1515

RESULT 11
MYS2_DICDI STANDARD; PRT; 2116 AA.
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN II HEAVY CHAIN, NON MUSCLE.
GN MICA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;

RN [1] SEQUENCE FROM N.A.
 RP MEDLINE-87092266; PubMed-3540939;
 RA Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
 RT "Conserved protein domains in a myosin heavy chain gene from
 Dictyostelium discoideum.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
 RN [2]
 RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
 RC STRAIN-AX2;
 RX MEDLINE-90353583; PubMed-2387408;
 RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
 Gerisch G.;
 RT "Replacement of threonine residues by serine and alanine in a
 phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
 RL FEBS Lett. 269:239-243(1990).
 RN [3]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE-88112226; PubMed-2828113;
 RA Wagle G., Noegel A., Scheel J., Gerisch G.;
 RT "Phosphorylation of threonine residues on cloned fragments of the
 Dictyostelium myosin heavy chain.";
 RL FEBS Lett. 227:71-75(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
 RX MEDLINE-95345066; PubMed-7619795;
 RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
 Rayment I.;
 RT "X-ray structures of the myosin motor domain of Dictyostelium
 discoideum complexed with MgADP.Befx and MgADP.ALf4.";
 RL Biochemistry 34:8960-8972(1995).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
 RX MEDLINE-95345067; PubMed-7619796;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
 truncated head of Dictyostelium discoideum myosin to 2.7-A
 resolution.";
 RL Biochemistry 34:8973-8981(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE-96206189; PubMed-8611530;
 RA Smith C.A., Rayment I.;
 RT "X-ray structure of the magnesium(II).ADP.vanadate complex of the
 Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
 RL Biochemistry 35:5404-5417(1996).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
 RX MEDLINE-97452580; PubMed-9305951;
 RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
 RT "X-ray structures of the MgADP, MgATPgammA, and MgAMPPNP complexes
 of the Dictyostelium discoideum myosin motor domain.";
 RL Biochemistry 36:11619-11628(1997).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
 RX MEDLINE-98070605; PubMed-9405148;
 RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I.;
 RT "X-ray crystal structure and solution fluorescence characterization
 of Mg.2(3')-O-(N-methylanthraniloyl) nucleotides bound to the
 Dictyostelium discoideum myosin motor domain.";
 RL J. Mol. Biol. 274:394-407(1997).
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -1- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHCK), 2 ALKALI
 LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 (MLC-2).
 CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CORTIX.
 CC -1- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMH) AND 1 HEAVY MEROMYOSIN (HMH). IT CAN BE FURTHER
 SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).

CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
 ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 POSITION (688).
 CC -1- SIMILARITY: THE GLOBULAR HEAD AA SEQUENCE SHOWS A HIGH DEGREE OF
 SIMILARITY WITH THE GLOBULAR HEAD SEQUENCES OF MUSCLE & NONMUSCLE
 HEAVY CHAINS. BY CONTRAST THE ROD SEQUENCE IS LESS CONSERVED, BUT
 THE PERIODICITIES OF HYDROPHOBIC & CHARGED RESIDUES, WHICH DICTATE
 THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE CONSERVED.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; MI4628; AAA33227.1; -;
 DR PIR; A26655; A26655.
 DR PIR; S00250; S00250.
 DR PDB; 1MMA; 03-DEC-97.
 DR PDB; 1MMD; 17-AUG-96.
 DR PDB; 1MMG; 03-DEC-97.
 DR PDB; 1MMN; 03-DEC-97.
 DR PDB; 1MND; 17-AUG-96.
 DR PDB; 1MNE; 17-AUG-96.
 DR PDB; 1VOM; 23-DEC-96.
 DR PDB; 1LVK; 28-JAN-98.
 DR Dictydb; DD01008; mhck.
 DR InterPro; IPR000048; -;
 DR InterPro; IPR001609; -;
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00663; myosin.head; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 KW Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
 KW Heptad repeat pattern; Methylation; Alkylation; Phosphorylation.
 FT DOMAIN 1 816 GLOBULAR HEAD (S1).
 FT DOMAIN 817 2116 RODLIKE TAIL (S2 AND LMN DOMAINS).
 FT NP_BIND 179 186 ATP.
 FT DOMAIN 638 660 ACTIN-BINDING.
 FT DOMAIN 738 752 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
 FT MOD_RES 678 678 ALKYLATION (SH-1).
 FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
 FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).
 FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
 SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770BB1EE56A1 CRC64;
 Query Match 7.1%; Score 145.5; DB:1; Length 2116;
 Best Local Similarity 23.1%; Pred. No. 2.1;
 Matches 82; Conservative 61; Mismatches 153; Indels 59; Gaps 12;
 QY 22 LKPLAGYVRSLSKQIEKNIITFNLDNLNLDILNSRLKRRKYFLDV-LESDLMOFKHISNEY 80
 DB 1673 IKSLVAEVDSEKQLEDEILA-----KDLVAKRALEVELEVRQLEEDRS 1723
 QY 81 IEDSFKLLNSEOKNILLKSKYIKESVENDIKFAQEGISYIEKVLAKYKDDLESIKV 140
 DB 1724 ELEDKSRRLTTEVEDI---KKYDAE-VEQNTKL-----DEAKKKLDDVDTKKQL 1771
 QY 141 KEKEKFPSSPPPTPPSPAKTDEQKSKFLPFLTNTLETLYNNLV-----NKIDVLYN 194
 DB 1772 EDEKKLINES-----EPAKKRLESENEDEFLAKDAEVKNSRAEKDRKYEKDLKD 1822
 QY 195 LKAKINDCNVEKDEAHVKITKLSDLKAIDDKIDLKFNWNTDFEAIKKLINDDTKKDMLGKL 254


```

Db 1823 TKYLNDEAATKTQTEGAALK-----EDIDELRSKLEOEQAKATQADSKKKTLEGE- 1875
Qy 255 LSTGLVQIFPNTII-----SKLIEGKFDMLNISQHCVKKQCPENSGGFRHLDRREE 307
Db 1876 IDNLRQALQEDSGKTKMRLEKEKRALEGELESELRETVEAEADSKSEAFQS---KRLVELEL 1932
Qy 308 CKLLNTKQEGDKCEENPNTPCNNNGCCDADATCTEE-----DSGSRKKITCE 357
Db 1933 EDARNLQKEIDAIEADAKSNLQREIVEAKGRLEESARTNSDRSRKRLAE 1987

RESULT 12
ECM_HUMAN
ID_1 ECM_HUMAN STANDARD; PRT; 1228 AA.
AC Q13201;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ENDOTHELIAL CELL MULTIMERIN PRECURSOR.
GN ECM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A., AND SEQUENCE OF 368-376.
RC TISSUP-Endothelial cells;
RX MEDLINE=95355440; PubMed=7629143;
RA Hayward C.P.M., Hassell J.A., Denomme G.A., Rachubinski R.A.,
RA Brown C., Kelton J.G.;
RT "The cDNA sequence of human endothelial cell multimerin. A unique
RT protein with RGDS, coiled-coil, and epidermal growth factor-like
RT domains and a carboxyl terminus similar to the globular domain of
RT complement C1q and collagen type VIII and X.";
RL J. Biol. Chem. 270:18246-18251(1995).
CC -1- FUNCTION: CARRIER PROTEIN FOR PLATELET (BUT NOT PLASMA) FACTOR
CC VWA. MAY PLAY A ROLE IN THE STORAGE AND STABILIZATION OF FACTOR V
CC IN PLATELETS.
CC -1- SUBUNIT: MULTIMERIC. COMPOSED OF VARYING SIZED, DISULFIDE-LINKED
CC MULTIMERS, THE SMALLEST OF WHICH IS A HOMOTRIMER. PROTEOLYSIS OF
CC THE PROMULTIMERIN IN THE N-TERMINAL REGION, LEADS TO THE MATURE
CC P155 FORM THAT IS STORED IN PLATELETS.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED BY ENDOTHELIAL CELLS AND
CC MEGAKARYOCYTES. STORED IN PLATELET ALPHA GRANULES AND ENDOTHELIAL
CC CELL WEIBEL-PALADE BODIES, FOLLOWING ACTIVATION OF THESE CELLS, IT
CC IS RELEASED AND ATTACHED TO MEGAKARYOCYTES, PLATELETS, ENDOTHELIUM
CC AND SUBENDOTHELIUM OF BLOOD VESSELS. NOT FOUND IN PLASMA. FOUND IN
CC VASCULAR TISSUES SUCH AS PLACENTA, LUNG, AND LIVER.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- PTM: EXTENSIVELY N-GLYCOSYLATED.
CC -1- DISEASE: DEFICIENCY IN MULTIMERIN DUE TO PROTEOLYTIC DEGRADATION
CC WITHIN THE PLATELET ALPHA GRANULES IS ASSOCIATED WITH AN AUTOSOMAL
CC DOMINANT BLEEDING DISORDER (FACTOR V QUEBEC).
CC -1- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U27109; AAC52065.1;
DR MIM; 601456;
DR InterPro; IPR000152;
DR InterPro; IPR000561;
DR InterPro; IPR001073;
DR Pfam; PF00008; EGF; 1.
DR PRINTS; PRO0007; COMPLEMENTC1Q.
DR PROSITE; PS01113; C1Q; 1.

```

```

DR PROSITE; PS00022; EGF_1; 1.
KW Signal; Glycoprotein; EGF-like domain; Coiled coil.
FT SIGNAL 19 POTENTIAL.
FT CHAIN 20 1228 ENDOTHELIAL CELL MULTIMERIN.
FT DOMAIN 1041 1077 EGF-LIKE.
FT DOMAIN 1119 1149 C1Q.
FT DOMAIN 333 365 COILED COIL (POTENTIAL).
FT DOMAIN 400 430 COILED COIL (POTENTIAL).
FT DOMAIN 503 523 COILED COIL (POTENTIAL).
FT DOMAIN 580 650 COILED COIL (POTENTIAL).
FT DOMAIN 675 726 COILED COIL (POTENTIAL).
FT DOMAIN 819 869 COILED COIL (POTENTIAL).
FT DOMAIN 189 192 POLY-SER.
FT DOMAIN 309 313 POLY-GLN.
FT SITE 186 188 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 1045 1056 BY SIMILARITY.
FT DISULFID 1050 1065 BY SIMILARITY.
FT DISULFID 1067 1076 BY SIMILARITY.
FT CARBOHYD 21 21 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 136 136 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 344 344 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 507 507 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 541 541 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 680 680 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 729 729 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 828 828 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 933 933 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 981 981 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 1020 1020 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 1075 1075 N-LINKED (GLCNAC. ) (POTENTIAL).
SQ SEQUENCE 1228 AA; 138071 MW; EAA98B0A17E2C4CD CRC64;

```

```

Query Match 6.5%; Score 133; DB 1; Length 1228;
Best Local Similarity 21.5%; Pred. No. 5.1;
Matches 71; Conservative 68; Mismatches 111; Indels 80; Gaps 17;

Qy 29 YRSLK-RQIEKNIITFNILNLDILNSRLKRRKYFLDVLDSLMQFKHSSNEYTIEDSFK 87
Db 330 YQAMKLTLLQKKIDNISITVNDVENT-----YSSLEGVSEDK---SREF--QSLLK 376
Qy 88 LLNSEQKNILKSYKYIKES---VENDIKFAQEGISYEVKVLAKYKDDLESIKKVIKEEK 144
Db 377 GLKSKSINVLIRD--IVREOFKIFQNDW---QETVAQLFKTVSSLSLESTROIQKVN 431
Qy 145 EKFPSSPTTPPSPAKTDEQKESKFL-----PFLNIETLYNNLYNKIDYILNLIKAK 198
Db 432 ESVVSI-----AAQKQFVLQVENRPTLDIVELRNHVNRQEMTLTCEKP 477
Qy 199 INDCNVEKDEAHVKITKLSDLKIDDDKIDLFKNNTDFEAKKIDNDTKKDMGLKLLSTG 258
Db 478 IKELEVK--QTHLE---GALEQHSRSILY-----YESLNKTLIS--KLKEVHEQLLSTE 524
Qy 259 LVQIFPNTIISKLIEGKFDMLNISQHCVKKQCPENSGGCFR--HLDE----- 304
Db 525 QVSDQKNAPAAESVSNVNTYMS-TLHENIKQSILMLQMFEDLHIQESKINNLTVSLEM 583
Qy 305 -----REEC-----KCLLNYKQEGDKCEEN 324
Db 584 EKESLRGECEDMLSKCRNDFKFLQKLDTEEN 613

```

LD 450 IN
QY 148 PS

```

DR EMBL; U28372; AAB64791.1; -.
DR PIR; S26710; S26710.
DR PIR; S34288; S34288.
DR SGD; S0002764; NUF1.
KW Coiled coil; Nuclear protein; Phosphorylation; Heptad repeat pattern.
FT DOMAIN 164 791 COILED COIL.
FT DOMAIN 54 59 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 726 731 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 742 747 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 731 944 ARG/TYR-RICH.
SQ SEQUENCE 944 AA; 111781 MW; 04FAA074B58A0BC8 CRC64;

Query Match 6.4%; Score 132; DB 1; Length 944;
Best Local Similarity 23.7%; Pred. No. 4.4;
Matches 85; Conservative 64; Mismatches 142; Indels 68; Gaps 18;

QY 27 GYRSLLKQIEKNTITNLDI-LNSRLKRRKYFDLVLESD--LMQFRHISSENYIIE 83
DB 381 GKLASLAQLTQ--LESKLNRDSQSLGSEELKKTNDKQKDIRIAREETVSKDERIID 438
QY 84 -----DSF--KLINSEKKNIL--LKSYYIKESVENDIKFAQEGISYYEKKVLAK 128
DB 439 LQKVKQLENDLFVKKTHSEKSTITNEUESDKLIKLENDLKVQAEKYSRMEKELKE 498
QY 129 YKDLESIKKVIKEEK----EKPPSPPTTPPPAKTDEQKESKFLPFLTNITLYNNL 184
DB 499 REFNYKISEKLEDEKTYLNEKISNLAENSQLNKKNEDNSTATHMK--ENYEKQLES 556
QY 185 VNKIDYLINK---AKINDCNVEKDEAHVKIT--KLSDKAIDDKI-DLFKN-----TND 234
DB 557 RKDIEEYKSAKDESEKIELKIRIAENSAKVSERKSKDIKQKDEQISDLTNLKLQDE 616
QY 235 FEATKLINDTTRKMDLGLKLLSTGLVQIFNPNTISKLIEGKFQDMLNISQHCVKKQCE 294
DB 617 ISSLKSII-DRYKKDF-----NQLASE--QSNIQHDLNLQILLENKL--- 656
QY 295 NSGCFRHLDERECLLNKQEGDKCEENPNPTCNENNGCCDADATCTEDSGSRKK 353
DB 657 -----IESEDELKSL-----RDSQKIEIE-NWKRKNYNNLSLENDRLITKESASDKR 703

RESULT 15
CIN8_YEAST
ID CIN8_YEAST STANDARD; PRT; 1038 AA.
AC P27895;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE KINESIN-LIKE PROTEIN CIN8.
GN CIN8 OR KSL2 OR YEL061C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=92317149; PubMed=1618897;
RA Hoyt M.A., He L., Loo K.K., Saunders W.S.;
RT "Two Saccharomyces cerevisiae kinesin-related gene products required for mitotic spindle assembly.";
RL J. Cell Biol. 118:109-120(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hymen R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;

```

```

Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 1021-1038 FROM N.A.
STRAIN=S288C;
RX MEDLINE=95172238; PubMed=7867803;
RA Rousselet G., Simon M., Ripoche P., Buhler J.M.;
RT "A second nitrogen permease regulator in Saccharomyces cerevisiae.";
RL FEBS Lett. 359:215-219(1995).
RN [4]
RP CHARACTERIZATION.
RC STRAIN=S288C;
RX MEDLINE=923354062; PubMed=1643659;
RA Saunders W.S., Hoyt M.A.;
RT "Kinesin-related proteins required for structural integrity of the mitotic spindle.";
RL Cell 70:451-458(1992).
CC -|- FUNCTION: REQUIRED FOR ASSEMBLY OF THE MITOTIC SPINDLE. INTERACT WITH SPINDLE MICROTUBULES TO PRODUCE AN OUTWARDLY DIRECTED FORCE ACTING UPON THE POLES. FOLLOWING SPINDLE ASSEMBLY, CIN8 AND KIP1 APPARENTLY ACT TO OPPOSE A FORCE THAT DRAWS SEPARATED POLES BACK TOGETHER. THIS FORCE SEEMS TO BE MEDIATED BY KAR3.
CC -|- SUBCELLULAR LOCATION: SPINDLE MICROTUBULES THAT LIE BETWEEN THE POLES.
CC -|- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11859; CAA77885.1; -.
DR EMBL; M90522; AAA34496.1; -.
DR EMBL; U18795; AAB65026.1; -.
DR EMBL; X79105; CAA55722.1; -.
DR PIR; B42641; B42641.
DR HSP; P17119; 3KAR.
DR SGD; S0000787; CIN8.
DR InterPro; IPR001752; -.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW Mitosis; Cell cycle.
KW DOMAIN 72 553 MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).
FT DOMAIN 554 657 COILED COIL (POTENTIAL).
FT DOMAIN 904 942 COILED COIL (POTENTIAL).
FT NP_BIND 166 173 ATP (BY SIMILARITY).
FT CONFLICT 254 254 D -> A (IN REF. 1).
FT CONFLICT 831 831 Q -> H (IN REF. 1).
SQ SEQUENCE 1038 AA; 117999 MW; 3A1FD7003EF89FBC CRC64;

Query Match 6.4%; Score 130.5; DB 1; Length 1038;
Best Local Similarity 25.3%; Pred. No. 5.7;
Matches 79; Conservative 54; Mismatches 126; Indels 53; Gaps 15;

QY 29 YRSLLKQIEKNTITNLDI-LNSRLKRRKYFDLVLESDLMQFKHISSE--YIEDSF 86
DB 508 YASKAKNIKKPKQLSGFTMKDIL---VNTWMLAKTKSDLLSTK---SKEGITWSQHY 561
QY 87 KLLNSEQKNILLKSYKIKESVENDIKFAQEGISYYEKKVLAKYKDDLESIKKVIKEEK 146
DB 562 KNLNSD-----LESYKKEVQCKREIE-----SITSKNALLVKDKLKS-KETIQSQNCQ 609
QY 147 FPSPPPTTPSPAKTDEQKESKFLPFLTNIE-TLYNNLVNKIDY-----LINKAKIND 201
DB 610 IESLTKTTHLRAQLDKQHK-----TETISDFNNKQLKLTVMQMALHDKYKRELD 661

```

